



HAL
open science

Multilevel Monte Carlo covariance estimation for the computation of Sobol' indices

Paul Mycek, Matthias de Lozzo

► **To cite this version:**

Paul Mycek, Matthias de Lozzo. Multilevel Monte Carlo covariance estimation for the computation of Sobol' indices. *SIAM/ASA Journal on Uncertainty Quantification*, 2019, 7 (4), pp. 1323-1348. 10.1137/18M1216389 . hal-01894503

HAL Id: hal-01894503

<https://hal.science/hal-01894503>

Submitted on 20 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Multilevel Monte Carlo covariance estimation for the computation of Sobol' indices

Paul Mycek* and Matthias De Lozzo†

Abstract. Crude and quasi Monte Carlo (MC) sampling techniques are common tools dedicated to estimating statistics (expectation, variance, covariance) of a random quantity of interest. We focus here on the uncertainty quantification framework where the quantity of interest is the output of a numerical simulator fed with uncertain input parameters. Then, sampling the output involves running the simulator for different samples of the inputs, which may be computationally time-consuming. To reduce the cost of sampling, a first approach consists in replacing the numerical simulator by a surrogate model that is cheaper to evaluate, thus making it possible to generate more samples of the output and therefore leading to a lower sampling error. However, this approach adds to the sampling error an unavoidable model error. Another approach, which does not introduce any model error, is the so-called multilevel MC (MLMC) method. Given a sequence of levels corresponding to numerical simulators with increasing accuracy and computational cost, MLMC combines samples obtained at different levels to construct an estimator at a reduced cost compared to standard MC sampling. In this paper, we derive and analyze multilevel covariance estimators and adapt the MLMC convergence theorem in terms of the corresponding covariances and fourth order moments. We propose a multilevel algorithm driven by a target cost as an alternative to typical algorithms driven by a target accuracy. These results are used in a sensitivity analysis context in order to derive a multilevel estimation of Sobol' indices, whose building blocks can be written as covariance terms in a pick-and-freeze formulation. These contributions are successfully tested on an initial value problem with random parameters.

Key words. Monte Carlo, Multilevel Monte Carlo, Parameter estimation, Covariance, Uncertainty quantification, Sensitivity analysis, Sobol' indices.

AMS subject classifications. 65C05, 65N55, 62L12, 62D05, 68Q25.

1. Introduction. Computational models are widely used to represent a quantity of interest by solving complex systems of equations that transcribe real life phenomena into mathematical language and that are discretized on dedicated meshes. Such simulators can take many numerical and explanatory variables as input parameters. They may be used in computer experimentation to explore the relationship between the variables and increase the knowledge of the real life phenomenon, while an accurate *in situ* experimentation would be too costly [47], even impossible in the case of forecasting purposes. However, these numerical simulators often have a large number of uncertain input parameters, which can lead to an important uncertainty over the model output [11, 36].

Commonly, in a probabilistic uncertainty quantification (UQ) framework, the uncertain input variables are modeled by random quantities (variables, vectors, fields) with prescribed probability distributions and propagated through the model. Their impact on the output are then quantified, usually by computing statistics such as their expectation, variance, prob-

*CERFACS, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France (mycek@cerfacs.fr).

†CECI, CNRS - CERFACS, 42 avenue Gaspard Coriolis, 31057 Toulouse Cedex 1, France. Current affiliation: IRT Saint-Exupéry, 3 rue Tarfaya, CS 34 436, 31405 Toulouse Cedex 4, France (matthias.delozzo@irt-saintexupery.com).

ability density function or threshold exceedance probability [17]. Sensitivity analysis (SA) methods [46] can be used to attribute shares of output randomness to individual input parameters or subsets thereof in interaction. The conclusions drawn SA may then help reduce the model output uncertainty by enhancing the characterization of the most influential input parameters and by using nominal deterministic values for the non-significant ones. Lastly, remaining input uncertainties can be controlled by data assimilation of real life measures facing the corresponding model output values [2].

Among the sensitivity measures used in SA studies, Sobol' indices [49] are particularly popular and powerful, and will be the focus of this paper. In UQ studies, such sensitivity indices are commonly estimated using Monte Carlo (MC) or quasi-Monte Carlo (QMC) sampling methods [44, 6, 40]. Precisely, the simulator is evaluated over a design of experiments whose elements are obtained from a set of independent realizations of the input random variables in the MC case or from a deterministic sequence of variables approaching such realizations efficiently in the QMC case. While the simplicity of these techniques is a significant advantage, the slow convergence of MC and QMC estimators with respect to the sample size represents a serious drawback when the simulator is computationally expensive.

A common way to address that limit is to compute the estimators from a cheap surrogate model parametrized by the statistical learning of a few evaluations of the simulator [15, 17]. Gaussian processes [43] and polynomial chaos (PC) expansions [39, 53] are examples of popular and widely-used surrogate models for UQ. In the case of PC surrogates, Sobol' indices can be directly retrieved from the expansion without requiring to re-sample the surrogate [50, 9]. One of the downsides of using a surrogate model, however, is that it introduces a model approximation error. In addition, constructing such a surrogate might be computationally expensive, especially when the number of uncertain input parameters is large. When the PC surrogate corresponds to the approximate solution of a stochastic problem sought in a tensor product space, as is the case in the stochastic Galerkin method [18, 39, 53], low-rank tensor techniques may be used to circumvent this so-called *curse of dimensionality* [42, 14, 12]. However, such formulations typically lead to dedicated methods for solving stochastic partial differential equations that do not allow for the non-intrusive use of the deterministic numerical simulator as a *black box*, which may be impractical when using, for instance, closed-source software.

When the accuracy of the numerical simulator and its associated computational cost can be adjusted, by changing the mesh resolution, for example, multilevel Monte Carlo (MLMC) sampling [28, 30, 29, 19, 20] can also be used to construct an estimator at a reduced cost compared to standard (Q)MC sampling. Early MLMC research in the late 1990s was conducted by Heinrich for the approximation of high dimensional parameter dependent integrals by MC sampling [28, 30, 29]. He developed for this purpose a multilevel variance reduction technique, popularized a decade later by Giles [19] who gave this method its modern name of MLMC. The main idea of MLMC consists in introducing a sequence of so-called levels, usually corresponding to a hierarchy of numerical simulators with increasing accuracy and associated cost of individual simulations. In practice, the different levels usually correspond to simulators having increasingly fine (spatial and/or temporal) mesh resolutions, so that simulators with a poor accuracy correspond to so-called "coarse" levels, while accurate simulators correspond to "fine" levels. The MLMC estimator is obtained by combining samples obtained from the

simulator at different levels in a specific way. In favorable cases, many samples come from cheap coarse levels in order to reduce the statistical error, while only few come from expensive fine levels to correct the resulting bias. This efficient combination leads to a reduced computational cost of estimating the statistics, in some cases with asymptotically the same cost as that of a single deterministic run of the simulator at the finest level.

From there, several extensions have appeared over the past decade, including methodological advances such as the combination of MLMC with QMC sampling [24], the use of antithetic variates [22, 23] and the possibility to define levels independently in each (space and/or time) direction, inspired from sparse grid ideas, leading to the so-called multiindex MC (MIMC) method [26]. Initially developed to estimate the expected solution of stochastic partial differential equations in computational finance [19, 21], MLMC has since been applied to various problems such as biological systems [41] or pollutant dispersion [35]. These techniques have also been developed in different areas of engineering and applied mathematics, for instance in reliability [1], for the estimation of covariance matrices in inverse problems [38], or for the construction of kriging metamodels in stochastic experiments [45]. A review of the most significant advances in the MLMC methodology, as well as codes and examples from different fields of application, can be found in [20].

Recent work has extended the MLMC estimate of expectation to the case of variance [3] and higher order central moments [4]. The case of covariance has also been addressed for the case of the covariance matrix used in the ensemble Kalman filter [32] for data assimilation. In this paper, we extend theorems of MLMC theory dedicated to estimating these statistics to the more general case of estimating arbitrary statistical parameters. We demonstrate in particular that covariance estimation fits in this framework by deriving upper bounds for the variance of the associated multilevel estimator. The resulting MLMC covariance estimator is then used in the context of global SA to estimate Sobol' indices, whose building blocks can be written as covariance terms in a pick-and-freeze formulation [34]. To our knowledge, this is the first contribution to the deployment of MLMC methods for variance-based sensitivity indices. We also devise an adaptive algorithm for choosing on the fly the optimal number of levels and sample sizes on each level, driven by a target overall computational budget. These theoretical and algorithmic contributions are successfully tested on an initial value problem with random parameters.

The paper is organized as follows. Section 2 presents the theoretical aspects by first describing the framework for the MLMC estimation of the expectation and of arbitrary statistics. Then our theoretical contributions to the MLMC estimation of the covariance are presented, as well as practical considerations leading to an adaptive algorithm for selecting on-the-fly optimal sample sizes on each level according to a given overall computational budget, rather than a target accuracy. Section 3 applies this MLMC estimator to the Sobol' indices and presents numerical results for an ordinary differential equation with random parameters. Finally, concluding remarks and prospects for further work are given in Section 4.

2. Multilevel Monte Carlo techniques for statistical parameter estimation. We consider an abstract numerical simulator described by the function:

$$(2.1) \quad \begin{aligned} f: \mathcal{X} &\rightarrow \mathbb{R} \\ \mathbf{x} &\mapsto f(\mathbf{x}) \equiv y, \end{aligned}$$

whose scalar input parameters $(x_1, \dots, x_d) = \mathbf{x} \in \mathcal{X} \subseteq \mathbb{R}^d$ are uncertain, leading to an uncertainty in the output value y . In a probabilistic uncertainty quantification framework, these uncertain parameters are commonly described by random variables defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The input vector \mathbf{x} is then replaced by a \mathcal{X} -valued random vector $\mathbf{X}: \Omega \rightarrow \mathcal{X}$ whose components are independent random variables with probability distributions given by expert knowledge; as a consequence, $Y \equiv f(\mathbf{X})$ is a random variable whose distribution is unknown. In this probabilistic context, we introduce the expectation operator \mathbb{E}

$$(2.2) \quad \mathbb{E}[Z] \equiv \int_{\Omega} Z(\omega) \, d\mathbb{P}(\omega),$$

for any random variable $Z: \Omega \rightarrow \mathbb{R}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$. Furthermore, let $L_p(\Omega, \mathbb{P})$, with $1 \leq p < \infty$, denote the space of random variables Z with finite L_p -norm

$$(2.3) \quad Z \in L_p(\Omega, \mathbb{P}) \iff \|Z\|_{L_p(\Omega, \mathbb{P})} \equiv \mathbb{E}[|Z|^p]^{1/p} < \infty.$$

Of particular interest is the space $L_2(\Omega, \mathbb{P})$ of so-called second-order random variables, which is equipped with the inner product $\langle \cdot, \cdot \rangle_{L_2(\Omega, \mathbb{P})}$ defined by:

$$(2.4) \quad \forall Z, Z' \in L_2(\Omega, \mathbb{P}), \quad \langle Z, Z' \rangle_{L_2(\Omega, \mathbb{P})} = \mathbb{E}[ZZ'] = \int_{\Omega} Z(\omega)Z'(\omega) \, d\mathbb{P}(\omega),$$

from which the L_2 -norm is induced, *i.e.* $\|Z\|_{L_2(\Omega, \mathbb{P})} = \langle Z, Z \rangle_{L_2(\Omega, \mathbb{P})}^{1/2}$.

In practice, the expectation of a random variable Z is approximated (estimated) by the following Monte Carlo (MC) estimator

$$(2.5) \quad E_M[Z] \equiv \frac{1}{M} \sum_{i=1}^M Z^{(i)},$$

where $\{Z^{(1)}, \dots, Z^{(M)}\}$ is an M -sample of Z , that is to say that $Z^{(1)}, \dots, Z^{(M)}$ are M independent and identically distributed (i.i.d.) replications of Z . The error of the MC estimator $E_M[Z]$, measured by the root mean square error (RMSE)

$$(2.6) \quad \text{RMSE}(E_M[Z], \mathbb{E}[Z]) \equiv \mathbb{E}[(E_M[Z] - \mathbb{E}[Z])^2]^{1/2},$$

is $\mathcal{O}(M^{-1/2})$, see, *e.g.*, [44]. Reducing this error by a factor of r thus implies increasing the sample size by a factor of r^2 . In practice, for the quantity of interest Y , obtaining a realization of the estimator $E_M[Y]$ requires M calls to the numerical simulator f , specifically at the elements of the M -sample $\{\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(M)}\}$ of \mathbf{X}

$$(2.7) \quad E_M[Y] = \frac{1}{M} \sum_{i=1}^M f(\mathbf{X}^{(i)}).$$

In uncertainty quantification, we are usually interested in the first central moments of Y and in quantiles or probabilities of threshold overrun [17]. Moreover, SA indices such as Sobol'

indices are commonly studied to quantify the shares of output variability attributable to the different input parameters [49]. Nevertheless, an accurate estimation of these statistical objects by standard MC sampling approaches requires a large number M of calls to the simulator f , namely $M = \mathcal{O}(\varepsilon^{-2})$ for an accuracy of $\mathcal{O}(\varepsilon)$, which is not always possible when the numerical simulator is computationally expensive and the CPU time budget is limited. When “coarser” simulators are available, for instance by degrading the mesh resolution of the original simulator, one can take advantage of the different levels of accuracy and associated computational cost by combining samples obtained from different levels in a multilevel variance reduction manner. This is the framework of MLMC sampling, which we now describe for the estimation of arbitrary statistical parameters.

2.1. MLMC estimator of the expectation. We first briefly recall the original MLMC estimator of the expectation [19]. Let $(f_\ell)_{\ell \geq 0}$ be a sequence of simulators with increasing accuracy and computational cost. In practice the different levels usually correspond to simulators having increasingly fine (spatial and/or temporal) mesh resolutions. For that reason, we will hereafter refer to small values of ℓ as “coarse” levels, while larger values of ℓ will be referred to as “fine” levels.

We denote by Y_ℓ the random variable $f_\ell(\mathbf{X})$, and we assume that $\lim_{\ell \rightarrow \infty} \mathbb{E}[Y_\ell] = \mathbb{E}[Y]$. The expectation $\mathbb{E}[Y_L]$ can be expressed as a telescoping sum

$$(2.8) \quad \mathbb{E}[Y_L] = \mathbb{E}[Y_0] + \sum_{\ell=1}^L \mathbb{E}[Y_\ell] - \mathbb{E}[Y_{\ell-1}],$$

which can be interpreted as a sum of corrections to an initial, coarse representation $\mathbb{E}[Y_0]$. For convenience, we let $Y_{-1} \equiv 0$ by convention and we write

$$(2.9) \quad \mathbb{E}[Y_L] = \sum_{\ell=0}^L \mathbb{E}[Y_\ell] - \mathbb{E}[Y_{\ell-1}].$$

Note that this telescoping sum converges to $\mathbb{E}[Y]$ as $L \rightarrow \infty$. Based on this rewriting, each expectation may be approximated by a standard MC estimator at the corresponding level of correction

$$(2.10) \quad E_L^{\text{ML}}[Y] = \sum_{\ell=0}^L E_{M_\ell}^{(\ell)}[Y_\ell] - E_{M_\ell}^{(\ell)}[Y_{\ell-1}],$$

where $E_{M_\ell}^{(\ell)}[Y_{\ell'}]$ is the standard MC estimator of $\mathbb{E}[Y_{\ell'}]$

$$(2.11) \quad E_{M_\ell}^{(\ell)}[Y_{\ell'}] = \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} Y_{\ell'}^{(\ell,i)}, \quad \text{with } Y_{\ell'}^{(\ell,i)} \equiv f_{\ell'}(\mathbf{X}^{(\ell,i)}), \quad \text{and } \ell' \in \{\ell-1, \ell\},$$

and where the random input vectors $\mathbf{X}^{(\ell,i)}$ are i.i.d. replications of \mathbf{X} . Replacing the single level MC estimators by their expression, Eq. (2.10) becomes

$$(2.12) \quad E_L^{\text{ML}}[Y] = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} [Y_\ell^{(\ell,i)} - Y_{\ell-1}^{(\ell,i)}] = \sum_{\ell=0}^L \frac{1}{M_\ell} \sum_{i=1}^{M_\ell} [f_\ell(\mathbf{X}^{(\ell,i)}) - f_{\ell-1}(\mathbf{X}^{(\ell,i)})],$$

highlighting the fact that the correction at each level ℓ is computed from the same input M_ℓ -sample $\{\mathbf{X}^{(\ell,i)}\}_{i=1,\dots,M_\ell}$, but using calls to different numerical simulators, f_ℓ and $f_{\ell-1}$.

2.2. MLMC estimator of arbitrary statistical parameters. The multilevel formulation described above, originally designed for the estimation of expected values, may be directly extended to the estimation of more general statistical parameters. Specifically, let θ be the parameter to be estimated, let $(\theta_\ell)_{\ell \geq 0}$ be a sequence of approximations of $\theta = \lim_{\ell \rightarrow \infty} \theta_\ell$. We then have

$$(2.13) \quad \theta_L = \sum_{\ell=0}^L T_\ell, \quad \text{with } T_\ell \equiv \theta_\ell - \theta_{\ell-1} \text{ and by convention } \theta_{-1} \equiv 0.$$

Similarly to the case of the expectation above, the MLMC estimator $\hat{\theta}_L^{\text{ML}}$ of θ may then be expressed as

$$(2.14) \quad \hat{\theta}_L^{\text{ML}} = \sum_{\ell=0}^L \hat{T}_{M_\ell}^{(\ell)},$$

where each $\hat{T}_{M_\ell}^{(\ell)}$ is a Monte Carlo estimator of T_ℓ using an M_ℓ -sample. The multilevel estimator $E_L^{\text{ML}}[Y]$ of the expectation $\mathbb{E}[Y]$, introduced above in Eq. (2.10), fits in this framework with $\hat{T}_{M_\ell}^{(\ell)} = E_{M_\ell}^{(\ell)}[Y_\ell] - E_{M_\ell}^{(\ell)}[Y_{\ell-1}]$.

We now introduce the bias of $\hat{\theta}$ as an estimator of a parameter θ as

$$(2.15) \quad \text{Bias}(\hat{\theta}, \theta) \equiv \mathbb{E}[\hat{\theta}] - \theta.$$

If $\text{Bias}(\hat{\theta}, \theta) = 0$, *i.e.* $\mathbb{E}[\hat{\theta}] = \theta$, we say that $\hat{\theta}$ is an *unbiased* estimator of θ . It is well-known that the mean square error (MSE) of $\hat{\theta}$ as an estimator of θ can be decomposed into a sum of two contributions, namely the variance of the estimator and its squared bias

$$(2.16) \quad \text{MSE}(\hat{\theta}, \theta) \equiv \mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{V}[\hat{\theta}] + \text{Bias}(\hat{\theta}, \theta)^2.$$

The root mean square error (RMSE) is simply defined as the square-root of the MSE, that is to say $\text{RMSE}(\hat{\theta}, \theta) \equiv \text{MSE}(\hat{\theta}, \theta)^{1/2}$. For the multilevel estimator $\hat{\theta}_L^{\text{ML}}$ defined in Eq. (2.14), it is easy to see that

$$(2.17) \quad \text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta) = \text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta_L) + (\theta_L - \theta),$$

where

$$(2.18) \quad \text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta_L) = \sum_{\ell=0}^L \text{Bias}(\hat{T}_{M_\ell}^{(\ell)}, T_\ell).$$

It is clear from that expression that if $\hat{T}_{M_\ell}^{(\ell)}$ is an unbiased estimator of T_ℓ on each level $\ell \leq L$, then $\text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta_L) = 0$. In that case, the total bias reduces to $\text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta) = \theta_L - \theta$ and thus the MSE becomes

$$(2.19) \quad \text{MSE}(\hat{\theta}_L^{\text{ML}}, \theta) = \mathbb{V}[\hat{\theta}_L^{\text{ML}}] + (\theta_L - \theta)^2.$$

In that case, the first term of the MSE, $\mathbb{V}[\hat{\theta}_L^{\text{ML}}]$, can be seen as pure *sampling error*, and the second term, $(\theta_L - \theta)^2$, can be seen as pure *discretization error*. It is important to note that the latter only involves the finest level L . Consequently, if the MSE remains large compared to the sampling error, an additional, finer level may need to be considered. Conversely, if the discretization error is small compared to the MSE, new multilevel simulations are required on the existing levels [19, 20]. Finally, we notice that if the estimators $\hat{T}_{M_\ell}^{(\ell)}$ are mutually uncorrelated, then $\mathbb{V}[\hat{\theta}_L^{\text{ML}}] = \sum_{\ell=0}^L \mathbb{V}[\hat{T}_{M_\ell}^{(\ell)}]$. In the sequel, we will solely consider estimators $\hat{T}_{M_\ell}^{(\ell)}$ that are mutually independent, and therefore uncorrelated.

Theorem 2.1 below is an essential theorem that ensures the convergence of the multilevel estimator at a bounded cost.

Theorem 2.1. *Let θ be a statistical parameter and $(\theta_\ell)_{\ell \geq 0}$ be a sequence of approximations of θ such that $\lim_{\ell \rightarrow \infty} \theta_\ell = \theta$. Let $(\hat{T}_{M_\ell}^{(\ell)})_{\ell \geq 0}$ be a sequence of mutually independent estimators, each $\hat{T}_{M_\ell}^{(\ell)}$ estimating $T_\ell \equiv \theta_\ell - \theta_{\ell-1}$ using an M_ℓ -sample, with $\theta_{-1} \equiv 0$ by convention. Let \mathcal{C}_ℓ denote the computational cost of evaluating a single member of the M_ℓ -sample used for the computation of $\hat{T}_{M_\ell}^{(\ell)}$, and let $(n_\ell)_{\ell \geq 0}$ be a sequence of positive integers satisfying $n_\ell \approx s^\ell$ for some fixed $s > 1$.*

Assume that there exists a fixed finite $m \in \mathbb{N}_0$ such that $\mathbb{V}[\hat{T}_{M_\ell}^{(\ell)}] \leq \mathcal{V}_\ell / (M_\ell - m)$, where \mathcal{V}_ℓ is independent of M_ℓ . Moreover, assume that there exist constants $\alpha, \beta, \gamma > 0$ such that, for any $\ell \geq 0$,

- (i) $|\text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta)| \lesssim n_\ell^{-\alpha}$,
- (ii) $\mathcal{V}_\ell \lesssim n_\ell^{-\beta}$,
- (iii) $\mathcal{C}_\ell \lesssim n_\ell^\gamma$.

Then for any tolerance $0 < \varepsilon < e^{-1}$, there exist a level $L \geq 0$ and a sequence of integers $(M_\ell)_{\ell=0}^L$, satisfying $M_\ell > m$ for all $0 \leq \ell \leq L$, such that $\text{RMSE}(\hat{\theta}_L^{\text{ML}}, \theta) \leq \varepsilon$ and

$$(2.20) \quad \text{Cost}_\varepsilon(\hat{\theta}_L^{\text{ML}}) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}} \mathbf{1}_{\min(\beta, \gamma) > 2\alpha} + \varepsilon^{-2} \left(\mathbf{1}_{\beta > \gamma} + (\log \varepsilon)^2 \mathbf{1}_{\beta = \gamma} + \varepsilon^{-\frac{\gamma - \beta}{\alpha}} \mathbf{1}_{\beta < \gamma} \right),$$

where $\text{Cost}_\varepsilon(\hat{\theta}_L^{\text{ML}})$ denotes the total cost of computing such a multilevel estimator.

In the theorem above, \lesssim and \approx have the same meaning as in [7, 3]. That is, for $a, b > 0$, $a \lesssim b$ means that a/b is bounded above by a constant independent of any parameters, and $a \approx b$ means that $a \lesssim b$ and $b \lesssim a$. Furthermore, with a slight abuse of the indicator function notation, for any logical proposition P , we define $\mathbf{1}_P$ as

$$(2.21) \quad \mathbf{1}_P = \begin{cases} 1 & \text{if } P \text{ is true,} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2.1 is an adaptation of the abstract convergence theorem introduced in [51, Theorem 4.1] to make it applicable to unbiased variance and covariance estimation (see sections 2.3 and 2.4 below). The main difference is that we require $\mathbb{V}[\hat{T}_{M_\ell}^{(\ell)}] \leq \mathcal{V}_\ell / (M_\ell - m)$ so that assumption (ii) implies $\mathbb{V}[\hat{T}_{M_\ell}^{(\ell)}] \lesssim (M_\ell - m)^{-1} n_\ell^{-\beta}$, while [51, Theorem 4.1], primarily designed

for the estimation of expected values, requires $\mathbb{V}[\hat{T}_{M_\ell}^{(\ell)}] \lesssim M_\ell^{-1/\delta} n_\ell^{-\beta}$, where $\delta = 1$ for MC and $\delta \in (1/2, 1]$ for QMC. The proof of Theorem 2.1 is along the same lines as that of [7, Theorem 1] and [51, Theorem 4.1].

Remark 2.2. When each $\hat{T}_{M_\ell}^{(\ell)}$ is an unbiased estimator of T_ℓ , assumption (i) in Theorem 2.1 can be relaxed and it is sufficient to have $|\theta_\ell - \theta_{\ell-1}| \lesssim n_\ell^{-\alpha}$ instead (see Appendix A).

Remark 2.3. It should be noted that there are cases where MLMC may fail or be ineffective. For instance, for coarse discretizations, the error may not be in the asymptotic regime, possibly preventing the assumptions of exponential bias and variance decay rate from holding for coarse levels. Furthermore, the geometry of the spatial computational domain is typically impacted by the discretization. For complex industrial problems, this may incur significant additional error and thus shift the asymptotic regime to finer discretizations. In such cases, one may opt for other methods (*e.g.* using a surrogate model) or design specific workarounds for MLMC.

2.3. MLMC covariance and variance estimator. In this work, we are particularly interested in estimating the covariance between two random variables (and, as a special case, the variance of a random variable). We define the MLMC estimator $C_L^{\text{ML}}[Y, Z]$ of the covariance $\mathbb{C}[Y, Z]$ between two second-order random variables Y and Z as

$$(2.22) \quad C_L^{\text{ML}}[Y, Z] = \sum_{\ell=0}^L C_{M_\ell}^{(\ell)}[Y_\ell, Z_\ell] - C_{M_\ell}^{(\ell)}[Y_{\ell-1}, Z_{\ell-1}],$$

where $C_{M_\ell}^{(\ell)}[Y_{\ell'}, Z_{\ell'}]$ is the single-level MC estimator

$$(2.23) \quad \begin{aligned} C_{M_\ell}^{(\ell)}[Y_{\ell'}, Z_{\ell'}] &= \frac{M_\ell}{M_\ell - 1} E_{M_\ell}^{(\ell)} \left[\left(Y_{\ell'} - E_{M_\ell}^{(\ell)}[Y_{\ell'}] \right) \left(Z_{\ell'} - E_{M_\ell}^{(\ell)}[Z_{\ell'}] \right) \right] \\ &= \frac{M_\ell}{M_\ell - 1} \left(E_{M_\ell}^{(\ell)}[Y_{\ell'} Z_{\ell'}] - E_{M_\ell}^{(\ell)}[Y_{\ell'}] E_{M_\ell}^{(\ell)}[Z_{\ell'}] \right). \end{aligned}$$

It can easily be seen that $C_{M_\ell}^{(\ell)}[Y_{\ell'}, Z_{\ell'}]$ is an *unbiased* estimator of $\mathbb{C}[Y_{\ell'}, Z_{\ell'}]$, implying that $\text{Bias}(C_L^{\text{ML}}[Y, Z], \mathbb{C}[Y, Z]) = \mathbb{C}[Y_L, Z_L] - \mathbb{C}[Y, Z]$. The multilevel estimator $V_L^{\text{ML}}[Y]$ of the variance $\mathbb{V}[Y] = \mathbb{C}[Y, Y]$ is obtained from the special case $Y = Z$

$$(2.24) \quad V_L^{\text{ML}}[Y] \equiv C_L^{\text{ML}}[Y, Y] = \sum_{\ell=0}^L V_{M_\ell}^{(\ell)}[Y_\ell] - V_{M_\ell}^{(\ell)}[Y_{\ell-1}], \quad V_{M_\ell}^{(\ell)}[Y_{\ell'}] \equiv C_{M_\ell}^{(\ell)}[Y_{\ell'}, Y_{\ell'}],$$

and, as a consequence, $\text{Bias}(V_L^{\text{ML}}[Y], \mathbb{V}[Y]) = \mathbb{V}[Y_L] - \mathbb{V}[Y]$.

Remark 2.4. For covariance estimation, sometimes it may not be possible — or desirable — to consider different levels for both random variables. This may be the case for instance when interested in the covariance between an input parameter and the output of a numerical simulator. In such a case, the multilevel estimator

$$(2.25) \quad \tilde{C}_L^{\text{ML}}[Y, Z] = \sum_{\ell=0}^L C_{M_\ell}^{(\ell)}[Y, Z_\ell] - C_{M_\ell}^{(\ell)}[Y, Z_{\ell-1}] = \sum_{\ell=0}^L C_{M_\ell}^{(\ell)}[Y, Z_\ell - Z_{\ell-1}]$$

can be used, where, by symmetry, the roles of Y and Z can be interchanged.

2.4. Variance of the multilevel estimators. Theorem 2.1 shows that it is essential to know an upper bound for the variance of the multilevel estimators. In this section, we first recall known expressions for the variance of some single-level MC estimators, then we use them to derive upper bounds for the variance of multilevel estimators.

For the sake clarity, we drop the unnecessary level-related sub- and superscripts for the single-level MC estimators. The variance of the MC estimators $E_M[Y]$, $C_M[Y, Z]$ and $V_M[Y]$ defined previously in Eq. (2.11), (2.23) and (2.24) can be expressed as

$$(2.26) \quad \forall Y \in L_2(\Omega, \mathbb{P}), \forall M > 0, \quad \mathbb{V}[E_M[Y]] = \mathbb{V}[Y]/M,$$

$$(2.27) \quad \forall Y, Z \in L_4(\Omega, \mathbb{P}), \quad \forall M > 1,$$

$$\mathbb{V}[C_M[Y, Z]] = \frac{\mathbb{M}^4[Y, Z]}{M} - \frac{(M-2)\mathbb{C}[Y, Z]^2 - \mathbb{V}[Y]\mathbb{V}[Z]}{M(M-1)},$$

$$(2.28) \quad \forall Y \in L_4(\Omega, \mathbb{P}), \quad \forall M > 1, \quad \mathbb{V}[V_M[Y]] = \frac{\mathbb{M}^4[Y]}{M} - \frac{M-3}{M(M-1)}\mathbb{V}[Y]^2,$$

where $\mathbb{M}^4[Y] \equiv \mathbb{E}[(Y - \mathbb{E}[Y])^4]$ and $\mathbb{M}^4[Y, Z] \equiv \mathbb{E}[(Y - \mathbb{E}[Y])^2(Z - \mathbb{E}[Z])^2]$. The expression for the expectation estimator is a standard result of estimation theory, while the expressions for the variance and covariance estimators can be found in [3, Corollary 4.2]. Note that Eq. (2.28) is simply a special case of Eq. (2.27), since $V_M[Y] = C_M[Y, Y]$. We can then derive the upper bounds

$$(2.29) \quad \mathbb{V}[C_M[Y, Z]] \leq \frac{1}{M-1} \sqrt{\mathbb{M}^4[Y] \mathbb{M}^4[Z]}, \quad \mathbb{V}[V_M[Y]] \leq \frac{1}{M-1} \mathbb{M}^4[Y],$$

for $M > 1$, see Appendix B. From these single-level results, we can derive upper bounds, or identity for the expectation, for the multilevel estimators defined in Eq. (2.10), (2.22) and (2.24):

$$(2.30) \quad \mathbb{V}[E_L^{\text{ML}}[Y]] = \sum_{\ell=0}^L \frac{1}{M_\ell} \mathbb{V}[\Delta_\ell^Y],$$

$$(2.31) \quad \mathbb{V}[C_L^{\text{ML}}[Y, Z]] \leq \frac{1}{2} \sum_{\ell=0}^L \frac{1}{M_\ell - 1} \left[\sqrt{\mathbb{M}^4[\Delta_\ell^Y] \mathbb{M}^4[\Sigma_\ell^Z]} + \sqrt{\mathbb{M}^4[\Delta_\ell^Z] \mathbb{M}^4[\Sigma_\ell^Y]} \right],$$

$$(2.32) \quad \mathbb{V}[V_L^{\text{ML}}[Y]] \leq \sum_{\ell=0}^L \frac{1}{M_\ell - 1} \sqrt{\mathbb{M}^4[\Delta_\ell^Y] \mathbb{M}^4[\Sigma_\ell^Y]},$$

where $\Delta_\ell^Y \equiv Y_\ell - Y_{\ell-1}$ and $\Sigma_\ell^Y \equiv Y_\ell + Y_{\ell-1}$. The expression for the expectation can be found in the original multilevel Monte Carlo paper [19], while the bound for the covariance estimator

is derived in Appendix C, following an idea similar to that for the variance estimator in [3, Theorem 5.1]. This derivation also bears resemblance with the analysis in [32]. The main idea relies on the assumption that the correction estimators $\hat{T}_{M_\ell}^{(\ell)}$ are mutually independent, so that $\mathbb{V}[C_L^{\text{ML}}[Y, Z]] = \sum_{\ell=0}^L \mathbb{V}[\hat{T}_{M_\ell}^{(\ell)}]$, and the rest of the proof thus essentially consists in bounding each individual $\mathbb{V}[\hat{T}_{M_\ell}^{(\ell)}]$ as required by Theorem 2.1, *i.e.* $\mathbb{V}[\hat{T}_{M_\ell}^{(\ell)}] \leq \mathcal{V}_\ell / (M_\ell - m)$. Table 1 summarizes main properties of the multilevel estimators derived above, with an additional property for $\tilde{C}_L^{\text{ML}}[Y, Z]$ that is straightforward to obtain.

Table 1: Summary of properties of the multilevel estimators.

Parameter	ML estimator	Correction at level ℓ	Variance contribution at level ℓ
θ	$\hat{\theta}_L^{\text{ML}}$	$\hat{T}_{M_\ell}^{(\ell)}$	\mathcal{V}_ℓ
$\mathbb{E}[Y]$	$E_L^{\text{ML}}[Y]$	$E_{M_\ell}^{(\ell)}[Y_\ell] - E_{M_\ell}^{(\ell)}[Y_{\ell-1}]$	$\mathbb{V}[\Delta_\ell^Y]$
$\mathbb{C}[Y, Z]$	$C_L^{\text{ML}}[Y, Z]$	$C_{M_\ell}^{(\ell)}[Y_\ell, Z_\ell] - C_{M_\ell}^{(\ell)}[Y_{\ell-1}, Z_{\ell-1}]$	$\frac{\sqrt{\mathbb{M}^4[\Delta_\ell^Y] \mathbb{M}^4[\Sigma_\ell^Z]} + \sqrt{\mathbb{M}^4[\Delta_\ell^Z] \mathbb{M}^4[\Sigma_\ell^Y]}}{2}$
$\mathbb{C}[Y, Z]$	$\tilde{C}_L^{\text{ML}}[Y, Z]$	$C_{M_\ell}^{(\ell)}[Y, Z_\ell] - C_{M_\ell}^{(\ell)}[Y, Z_{\ell-1}]$	$\sqrt{\mathbb{M}^4[Y] \mathbb{M}^4[\Delta_\ell^Z]}$
$\mathbb{V}[Y]$	$V_L^{\text{ML}}[Y]$	$V_{M_\ell}^{(\ell)}[Y_\ell] - V_{M_\ell}^{(\ell)}[Y_{\ell-1}]$	$\sqrt{\mathbb{M}^4[\Delta_\ell^Y] \mathbb{M}^4[\Sigma_\ell^Y]}$

From these results, we can derive Theorem 2.5 for the multilevel covariance estimator as a special case of Theorem 2.1.

Theorem 2.5. *Let Y and Z be random variables, and let Y_ℓ and Z_ℓ denote their respective approximations at level ℓ . Let \mathcal{C}_ℓ be the cost of evaluating a single member of the M_ℓ -sample used for the computation of $C_{M_\ell}^{(\ell)}[Y_\ell, Z_\ell] - C_{M_\ell}^{(\ell)}[Y_{\ell-1}, Z_{\ell-1}]$, and let $(n_\ell)_{\ell \geq 0}$ be an exponentially increasing sequence of positive integers satisfying $n_\ell \approx s^\ell$ for some fixed $s > 1$. Moreover, assume that $\{\mathbb{M}^4[Y_\ell]\}_{\ell \geq 0}$ and $\{\mathbb{M}^4[Z_\ell]\}_{\ell \geq 0}$ are uniformly bounded and that there exist constants $\alpha, \beta, \gamma > 0$ such that, for any $\ell \geq 0$,*

- (i) $|\mathbb{C}[Y_\ell, Z_\ell] - \mathbb{C}[Y_{\ell-1}, Z_{\ell-1}]| \lesssim n_\ell^{-\alpha}$,
- (ii) $\mathbb{M}^4[\Delta_\ell^Y]^{1/2} + \mathbb{M}^4[\Delta_\ell^Z]^{1/2} \lesssim n_\ell^{-\beta}$,
- (iii) $\mathcal{C}_\ell \lesssim n_\ell^\gamma$.

Then for any tolerance $0 < \varepsilon < e^{-1}$, there exist an integer $L \geq 0$ and a sequence of integers $(M_\ell)_{\ell=0}^L$, satisfying $M_\ell > 1$ for all $0 \leq \ell \leq L$, such that

$$(2.33) \quad \text{RMSE}(C_L^{\text{ML}}[Y, Z], \mathbb{C}[Y, Z]) \leq \varepsilon$$

and

$$(2.34) \quad \text{Cost}_\varepsilon(C_L^{\text{ML}}[Y, Z]) \lesssim \varepsilon^{-\frac{\gamma}{\alpha}} \mathbf{1}_{\min(\beta, \gamma) > 2\alpha} + \varepsilon^{-2} \left(\mathbf{1}_{\beta > \gamma} + (\log \varepsilon)^2 \mathbf{1}_{\beta = \gamma} + \varepsilon^{-\frac{\gamma - \beta}{\alpha}} \mathbf{1}_{\beta < \gamma} \right).$$

Assumption (i) results from the fact that the level corrections $C_{M_\ell}^{(\ell)}[Y_\ell, Z_\ell] - C_{M_\ell}^{(\ell)}[Y_{\ell-1}, Z_{\ell-1}]$ are unbiased estimators of $\mathbb{C}[Y_\ell, Z_\ell] - \mathbb{C}[Y_{\ell-1}, Z_{\ell-1}]$, so the relaxed assumption can be used (see Remark 2.2). Furthermore, if $\{\mathbb{M}^4[Y_\ell]\}_{\ell \geq 0}$ and $\{\mathbb{M}^4[Z_\ell]\}_{\ell \geq 0}$ are uniformly bounded, then so are $\{\mathbb{M}^4[\Sigma_\ell^Y]\}_{\ell \geq 0}$ and $\{\mathbb{M}^4[\Sigma_\ell^Z]\}_{\ell \geq 0}$, and thus assumption (ii) of Theorem 2.1 holds (see Appendix C).

Remark 2.6. It should be stressed that the bounds and the corresponding MLMC theorems can be adapted to the case of random fields, using appropriate spaces and norms as described in [3] for the expectation and variance MLMC estimators.

2.5. Adaptive MLMC algorithms. Given a sequence of discretization levels described, *e.g.*, by the number of grids points n_ℓ on each level ℓ , the choice of L and $(M_\ell)_{\ell \leq L}$ in Theorem 2.1 depends on the inequality constants in assumptions (i)–(iii). We extend here the adaptive algorithm designed in [19, 7, 20] to estimate on the fly the optimal number of levels L and samples M_ℓ on each level.

First, we consider that \hat{T}_ℓ is an unbiased estimator of T_ℓ on each level and that the assumptions of Theorem 2.1 hold. Second, we remark that in order to achieve a RMSE smaller than or equal to some target ε , it is sufficient to have, for some given $\rho \in (0, 1)$,

$$(2.35) \quad |\text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta)| \leq \rho^{1/2} \varepsilon \equiv \varepsilon_b, \quad \text{and } \mathcal{V} \leq (1 - \rho) \varepsilon^2 \equiv \varepsilon_v^2,$$

where $\mathcal{V} \equiv \sum_{\ell \leq L} \mathcal{V}_\ell / (M_\ell - m)$. From these considerations, we propose adaptive algorithms to select the optimal values of L and $(M_\ell)_{\ell \geq 0}$ where the objective function can be either the accuracy of the MLMC estimator or its computational cost.

2.5.1. Target accuracy approach. Following an approach similar to that of [7, 20], we determine that the number of samples M_ℓ on each level required to achieve some target accuracy $\mathcal{V} = \varepsilon_v^2$ for a minimal total cost $\mathcal{C} = \sum_{\ell \leq L} M_\ell \mathcal{C}_\ell$ is given by:

$$(2.36) \quad \forall \ell \leq L, \quad M_\ell = m + 1 + \left\lceil \varepsilon_v^{-2} \mathcal{S}_L \sqrt{\mathcal{V}_\ell / \mathcal{C}_\ell} \right\rceil > m, \quad \mathcal{S}_L \equiv \sum_{\ell \leq L} \sqrt{\mathcal{V}_\ell \mathcal{C}_\ell},$$

where $\lfloor \cdot \rfloor$ denotes the floor function. This naturally leads to the adaptive procedure described in Algorithm 1, adapted from [20, Algorithm 1].

It should be stressed that this algorithm relies on the additional assumptions that

$$(2.37) \quad \forall \ell > 1, \quad d_\ell \equiv \theta_\ell - \theta_{\ell-1} = s^{-\alpha} d_{\ell-1} \quad \text{and } \mathcal{V}_\ell = s^{-\beta} \mathcal{V}_{\ell-1}.$$

The first identity implies that

$$(2.38) \quad \text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta) = \theta_L - \theta = - \sum_{\ell > L} d_\ell = -(s^\alpha - 1)^{-1} d_L,$$

where d_L can be approximated on the fly using available samples, see [20, Section 3.1].

Roughly speaking, the algorithm starts with a small number of coarse levels, *e.g.* 3, and a small number of samples on each level, from which the variance contributions \mathcal{V}_ℓ are estimated in order to compute the optimal sample sizes M_ℓ using Eq. (2.36). Note that these sample

Algorithm 1: Simplified MLMC algorithm for target accuracy strategy, adapted from [20, Algorithm 1].

```

1 Choose  $L_{\min} \geq 2$ ,  $M_{\text{init}} > m$ ,  $\varepsilon < e^{-1}$  and  $\rho \in (0, 1)$ ;
2 Set  $L = L_{\min}$  and  $M_\ell = M_{\text{init}}$  samples on levels  $\ell \leq L$ ;
3 while extra samples need to be evaluated do
4   evaluate extra samples on each level  $0 \leq \ell \leq L$ ;
5   compute/update estimates for  $d_\ell$  and  $\mathcal{V}_\ell$  from samples on levels  $\ell \leq L$ ;
6   update optimal number of samples  $M_\ell$  using Eq. (2.36) on levels  $\ell \leq L$ ;
7   if  $(s^\alpha - 1)^{-1}|d_L| > \varepsilon_b$  then
8      $L \leftarrow L + 1$ ;
9     initialize  $\mathcal{V}_L$  from  $\mathcal{V}_{L-1}$  using Eq. (2.37);
10    define optimal number of samples  $M_L$  using Eq. (2.36);
11  end if
12 end while

```

sizes may be under- or over-estimated due to the limited initial number of samples, so in practice, this process is repeated until the estimated optimal sample sizes stop increasing. This guarantees that $\mathcal{V} \leq \varepsilon_v^2$. Then the bias is computed using Eq. (2.38). If it satisfies $|\text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta)| \leq \varepsilon_b$, then the RMSE achieves the prescribed tolerance ε , and the algorithm can stop. Otherwise, a new level needs to be created. The second identity in Eq. (2.37) is used to estimate \mathcal{V}_L from \mathcal{V}_{L-1} for a newly created level L , then Eq. (2.36) is used to estimate the optimal number of samples on this new level. The overall process is repeated until the bias is sufficiently small. More details regarding the actual implementation of such an algorithm can be found in [20, Section 3.4].

Remark 2.7. Note that more involved strategies have been proposed in the literature. For instance, in [8] the authors propose a continuation MLMC (CMLMC) algorithm that iteratively applies MLMC with a sequence of decreasing tolerances, until the final target accuracy is reached. The number of levels is updated at every iteration, as well as the splitting parameter ρ (see Eq. (2.35)), and the variance contribution \mathcal{V}_ℓ is estimated using a Bayesian setting. In addition, the authors exploit the asymptotic normality of the MLMC expectation estimator to enforce the tolerance criterion on the statistical error with a confidence level $1 - \delta$, for some prescribed failure probability $\delta \in (0, 1)$. This iterative CMLMC approach computes increasingly accurate estimates of the bias and statistical error, which directly impact the estimation of the optimal number of samples per level, as well as the creation of new levels. In [27], the authors also discuss optimal mesh hierarchies and optimal error splitting.

2.5.2. Target cost approach. From a practical point of view, it may not be easy to prescribe a target accuracy. On the contrary, one may want to prescribe a target cost depending on a given available CPU budget. Using a technique similar to that proposed in Section 2.5.1, we find that the number of samples required to achieve some target cost $\mathcal{C} = \eta > 0$ with a

minimal total $\mathcal{V} \equiv \sum_{\ell \leq L} \mathcal{V}_\ell / (M_\ell - m)$, is given by

$$(2.39) \quad \forall \ell \leq L, \quad M_\ell = m + 1 + \left\lceil (\tilde{\eta}_L / \mathcal{S}_L) \sqrt{\mathcal{V}_\ell / \mathcal{C}_\ell} \right\rceil > m, \quad \tilde{\eta}_L \equiv \eta - m \sum_{\ell' \leq L} \mathcal{C}_{\ell'}.$$

Based on this expression, a similar algorithm to that for a target accuracy could be used. However, whenever a new level needs to be created, the resulting additional samples on the newly created level would cause the target cost to be exceeded. So instead, we resort to an algorithm inspired by that proposed in [24, Section 5], [20, Algorithm 2] and [37] for multilevel QMC. Starting from a small initial number of levels and initial number of samples on each level, it selects the ‘‘optimal’’ level on which to increase the sample size by a factor $\tau > 1$. Specifically, it selects the level ℓ^* for which the ratio between the variance reduction and the additional computational effort is maximal,

$$(2.40) \quad \ell^* = \arg \max_{0 \leq \ell \leq L} \frac{\mathcal{V}_\ell}{(M_\ell - m)(\tau M_\ell - m)\mathcal{C}_\ell}.$$

Whenever the overall variance becomes too small compared to the squared bias in the sense of Eq. (2.35), specifically whenever $\rho \mathcal{V} \leq (1 - \rho) \text{Bias}(\hat{\theta}_L^{\text{ML}}, \theta)^2$, a new (finer) level is created. The approach is outlined in Algorithm 2. Similarly to the CMLMC algorithm mentioned previously, the factor $\tau > 1$ allows for a moderate growth of the sample sizes during the early iterations of the algorithm, thus preventing poor estimations of \mathcal{V}_ℓ from inducing unnecessarily large increases on fine levels. A more involved version of the algorithm could allow for τ to take different values τ_ℓ depending on the level, typically larger values on coarse levels and smaller, more conservative values on fine levels.

Algorithm 2: Simplified MLMC algorithm for target cost strategy, inspired by [24] and [20, Algorithm 2].

- 1 Choose $L_{\min} \geq 2$, $M_{\text{init}} > m$, $\rho \in (0, 1)$, $\tau > 1$;
 - 2 Set $L = L_{\min}$, $\mathcal{C} = 0$ and $M_\ell = M_{\text{init}}$ samples on levels $\ell \leq L$;
 - 3 **while** $\mathcal{C} \leq \eta$ **do**
 - 4 evaluate extra samples on each level $\ell \leq L$ and update $\mathcal{C} = \sum_{\ell \leq L} M_\ell \mathcal{C}_\ell$;
 - 5 compute/update estimates for d_ℓ , \mathcal{V}_ℓ and \mathcal{V} from samples on levels $\ell \leq L$;
 - 6 **if** $\rho(s^\alpha - 1)^2 \mathcal{V} > (1 - \rho)d_L^2$ **then**
 - 7 select level ℓ^* based on Eq. (2.40);
 - 8 $M_{\ell^*} \leftarrow \lceil \tau M_{\ell^*} \rceil$;
 - 9 **else**
 - 10 $L \leftarrow L + 1$;
 - 11 $M_L \leftarrow M_{\text{init}}$;
 - 12 **end if**
 - 13 **end while**
-

Algorithm 2 differs from that proposed in [24, 20, 37] in that it is designed to comply with a given computational budget, as opposed to seeking to achieve a target accuracy. Another minor difference is the way the optimal level ℓ^* is chosen (see Eq. (2.40)).

3. Application to the MLMC estimation of Sobol' indices. When the input parameters of a numerical simulator f are subject to uncertainties, two important questions arise: “how uncertain is the simulator output?” and “how do the different input uncertainties contribute to this output uncertainty?”. Both issues occur in the framework of uncertainty quantification and risk management. The second one corresponds to the engineering branch called *sensitivity analysis*, whose objective is to highlight the contributions of the different uncertain input parameters X_1, \dots, X_d of the model f to the output $Y = f(\mathbf{X})$; see, *e.g.*, [46]. In this section, we focus on Sobol' indices, which are global sensitivity measures based on a decomposition of the output variance. Specifically, we demonstrate how to apply the MLMC methodology for covariance estimation to the computation of such indices, in order to reduce the overall cost of their estimation.

3.1. Sobol' indices. Under the assumption that $f(\mathbf{X})$ is square-integrable, *i.e.* $\mathbb{E}[f^2(\mathbf{X})] < \infty$, and that the random variables X_1, X_2, \dots, X_d are independent, there is a unique orthogonal output decomposition [31, 25, 13, 49, 52]

$$(3.1) \quad f(\mathbf{X}) = \sum_{\emptyset \subseteq u \subseteq I_d} f_u(\mathbf{X}_u),$$

with the iterative construction

$$(3.2) \quad \begin{cases} f_\emptyset(\mathbf{X}_\emptyset) \equiv f_\emptyset \equiv \mathbb{E}[f(\mathbf{X})], \\ f_u(\mathbf{X}_u) \equiv \mathbb{E}_{\bar{u}}[f(\mathbf{X})|\mathbf{X}_u] - \sum_{\emptyset \subseteq v \subsetneq u} f_v(\mathbf{X}_v), \quad \forall \emptyset \subsetneq u \subseteq I_d, \end{cases}$$

where $I_d \equiv \{1, \dots, d\}$, $\mathbf{X}_u \equiv \{X_i\}_{i \in u}$ for any $\emptyset \subsetneq u \subseteq I_d$, and $\bar{u} \equiv I_d \setminus u$. For clarity, the subscript in the expectation notation (and subsequently in the variance notation) indicates the indices of the variables over which the integration is performed. The subscript is omitted whenever the integration is performed over all the variables.

Then, it is straightforward to decompose the output variance $\mathbb{V}[f(\mathbf{X})]$ with respect to the different combinations of input parameters \mathbf{X}_u and to analyze the different shares of variance. Precisely, the Sobol' index associated to \mathbf{X}_u is defined as the share of output variance attributable to \mathbf{X}_u individually [49]:

$$(3.3) \quad S_u = \frac{\mathbb{V}_u[f_u(\mathbf{X}_u)]}{\mathbb{V}[f(\mathbf{X})]} = \frac{V_u}{\mathbb{V}[f(\mathbf{X})]} - \sum_{\emptyset \subseteq v \subsetneq u} S_v, \quad V_u \equiv \mathbb{V}_u[\mathbb{E}_{\bar{u}}[f(\mathbf{X})|\mathbf{X}_u]],$$

for any $\emptyset \subsetneq u \subseteq I_d$. Similarly, the total-order Sobol' index associated to \mathbf{X}_u is defined as the share of output variance attributable to \mathbf{X}_u individually and in interaction:

$$(3.4) \quad S_u^T = \sum_{u \subseteq v \subseteq I_d} S_v = 1 - \frac{V_{\bar{u}}}{\mathbb{V}[f(\mathbf{X})]}.$$

As mentioned in the introduction, Sobol' indices may be estimated by direct MC or QMC sampling of the numerical simulator [44, 6, 40]. Alternatively, a surrogate model (*e.g.* a Gaussian process [43] or a PC surrogate [39, 53, 50, 9]) may first be constructed to reduce the

sampling cost. In this paper, we focus on applying the MLMC methodology to the estimation of Sobol' indices. In particular, we are interested in the estimation of V_u , which is at the basis of the construction of both S_u and S_u^T , and which may be rewritten as a covariance term, leading to the so-called pick-and-freeze formulation [34] (see [33, Section 3.1.2, Lemma 1] for the proof)

$$(3.5) \quad V_u \equiv \mathbb{V}_u[\mathbb{E}_{\bar{u}}[f(\mathbf{X})|\mathbf{X}_u]] = \mathbb{C}[f(\mathbf{X}_u \oplus \mathbf{X}_{\bar{u}}), f(\mathbf{X}_u \oplus \mathbf{X}'_{\bar{u}})] = \mathbb{C}[f(\mathbf{X}), f(\mathbf{X}_u \oplus \mathbf{X}'_{\bar{u}})],$$

where

- $\mathbf{X}_u \oplus \mathbf{X}'_{\bar{u}}$ denotes the \mathbb{R}^d -valued random vector whose i th component is X_i if $i \in u$ and X'_i otherwise, where X_i and X'_i are i.i.d.; and
- $\mathbf{X}_u \oplus \mathbf{X}_{\bar{u}} = \mathbf{X}$.

In other words, V_u represents the covariance between the outputs $Y \equiv f(\mathbf{X}_u \oplus \mathbf{X}_{\bar{u}})$ and $Y_u \equiv f(\mathbf{X}_u \oplus \mathbf{X}'_{\bar{u}})$ where the second output is obtained by “freezing” the inputs \mathbf{X}_u corresponding to parameters indexed by $i \in u$, and considering an i.i.d. “copy” of the other parameters, hence the term “pick-and-freeze.”

Remark 3.1. This formulation also shows that Sobol' indices can be expressed in terms of correlation coefficients, specifically

$$(3.6) \quad \forall \emptyset \subsetneq u \subseteq I_d, \quad S_u = \rho_{Y, Y_u} - \sum_{\emptyset \subsetneq v \subsetneq u} S_v,$$

where $\rho_{Y, Z} \equiv \frac{\mathbb{C}[Y, Z]}{\sqrt{\mathbb{V}[Y] \mathbb{V}[Z]}}$ denotes the Pearson correlation coefficient between the random variables Y and Z . In particular, first-order Sobol' indices may be defined as

$$(3.7) \quad \forall i \in I_d, \quad S_i \equiv S_{\{i\}} = \rho_{Y, Y_i}.$$

Remark 3.2. It should be stressed that the terms in Eq. (3.3), *i.e.* V_u , $\mathbb{V}[f(\mathbf{X})]$ and $\sum_{\emptyset \subsetneq v \subsetneq u} S_v$ are estimated independently. Here we focus on the estimation of the covariance term V_u . The denominator $\mathbb{V}[f(\mathbf{X})]$ is a normalizing constant so that the indices can be thought of as shares of the total output variance. This constant may be estimated once and for all since it does not depend on u . It may also be omitted when one simply seeks to compare the contributions of (groups of) inputs, in which case one would typically be interested in estimating

$$(3.8) \quad \mathbb{V}_u[f_u(\mathbf{X}_u)] = V_u - \sum_{\emptyset \subsetneq v \subsetneq u} \mathbb{V}_v[f_v(\mathbf{X}_v)].$$

The last term, $\sum_{\emptyset \subsetneq v \subsetneq u} S_v$, arises from the iterative construction. When estimating S_u , the estimates of lower-order indices S_v , already available (since $v \subsetneq u$), are reused. It is clear from this construction that the estimates of Sobol' indices accumulate error from the estimation of the lower-order indices. This is nonetheless how Sobol' indices are typically estimated in MC-like sampling methods (plain MC, QMC), hence the focus on V_u .

3.2. Numerical experiments. We consider the following initial value problem with growth coefficient $\lambda \in \mathbb{R}$ and initial condition $u_0 \in \mathbb{R}$

$$(3.9) \quad \begin{cases} \frac{du}{dt}(t) = \lambda u(t), & t \in (0, 1], \\ u(0) = u_0, \end{cases}$$

whose solution is given by $u(t) = u_0 e^{\lambda t}$. We define the abstract function F that maps a pair of parameter values $(u_0, \lambda) \in \mathbb{R}^2$ of the input parameters and a time $t \in [0, 1]$ to the solution of (3.9) at time t with those parameters

$$(3.10) \quad \begin{aligned} F: \mathbb{R}^2 \times [0, 1] &\rightarrow \mathbb{R} \\ (u_0, \lambda, t) &\mapsto u_0 e^{\lambda t}. \end{aligned}$$

Let us now consider the input parameters as uncertain and investigate the sensitivity of the output to the uncertain inputs. For any $t \in (0, 1]$, we have (see Appendix D)

$$(3.11) \quad V_{U_0}(t) \equiv \mathbb{C}[F(U_0, \Lambda, t), F(U_0, \Lambda', t)] = \mathbb{E}[e^{\Lambda t}]^2 \mathbb{V}[U_0],$$

$$(3.12) \quad V_{\Lambda}(t) \equiv \mathbb{C}[F(U_0, \Lambda, t), F(U_0', \Lambda, t)] = \mathbb{E}[U_0]^2 \mathbb{V}[e^{\Lambda t}],$$

$$(3.13) \quad V(t) \equiv \mathbb{V}[F(U_0, \Lambda, t)] = \mathbb{E}[U_0^2] \mathbb{E}[e^{2\Lambda t}] - \mathbb{E}[U_0]^2 \mathbb{E}[e^{\Lambda t}]^2,$$

where U_0 and Λ are independent random variables representing the input uncertainty and U_0' and Λ' are i.i.d. copies of U_0 and Λ , respectively. In addition, letting $\Lambda \sim \mathcal{N}(\mu, \sigma^2)$ and $U_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$, we obtain the following first-order Sobol' indices

$$(3.14) \quad S_{U_0}(t) = \frac{\sigma_0^2 e^{-\sigma^2 t^2}}{\sigma_0^2 + \mu_0^2 (1 - e^{-\sigma^2 t^2})}, \quad S_{\Lambda}(t) = \frac{\mu_0^2 (1 - e^{-\sigma^2 t^2})}{\sigma_0^2 + \mu_0^2 (1 - e^{-\sigma^2 t^2})},$$

as well as the Sobol' index for the interaction effects between U_0 and Λ :

$$(3.15) \quad S_{U_0, \Lambda}(t) = \frac{\sigma_0^2 (1 - e^{-\sigma^2 t^2})}{\sigma_0^2 + \mu_0^2 (1 - e^{-\sigma^2 t^2})}.$$

Moreover, the total variance is given by

$$(3.16) \quad V(t) = e^{2\mu t + \sigma^2 t^2} [\sigma_0^2 e^{\sigma^2 t^2} + \mu_0^2 (e^{\sigma^2 t^2} - 1)].$$

Finally, in what follows, we consider the following hyper-parameters for the distributions of U_0 and Λ :

$$(3.17) \quad \mu = -1, \quad \sigma = 0.25, \quad \mu_0 = 10, \quad \sigma_0 = 2,$$

which correspond to the values given in [48, Example 10.10] for the same problem. The Sobol' indices with these particular hyper-parameters are plotted as a function of time in Fig. 1. At the initial time, the growth coefficient λ is absent from the solution of the differential equation

(3.9); only the initial condition u_0 intervenes. The result is a first-order Sobol' index equal to 1 for the random variable U_0 which totally explains the variance of $F(U_0, \Lambda, 0)$ and a Sobol' index equal to 0 for Λ . This is reflected in the formulas (3.14) as well as in Fig. 1. Furthermore, as $t \rightarrow \infty$, the first-order Sobol' index for Λ increases and tends to $25/26 \approx 0.96$ while that for U_0 decreases and tends to 0, and that for the interaction between U_0 and Λ increases from 0 and tends to $1/26 \approx 0.04$. Lastly, U_0 and Λ contribute in the same way to the variance of $F(U_0, \Lambda, t)$ when $t = \sigma^{-1} \sqrt{\ln(1 + \sigma_0^2 \mu_0^{-2})} = 4\sqrt{\ln(1.04)} \approx 0.79$.

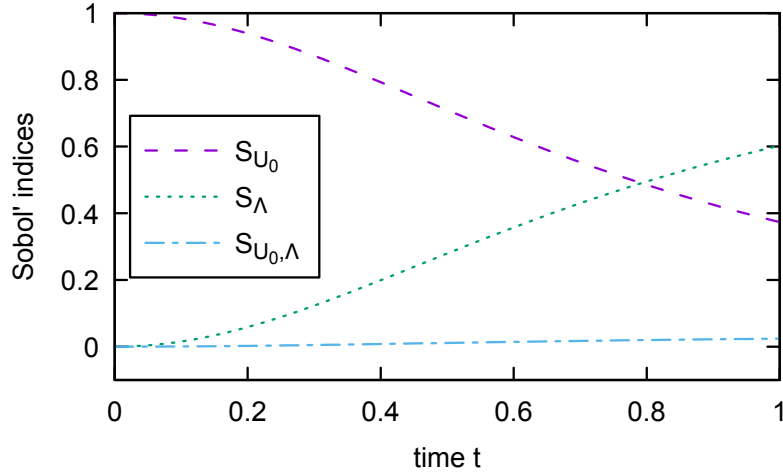


Figure 1: Evolution of the Sobol' indices defined in Eq. (3.14) and (3.15) as a function of time t , with $\Lambda \sim \mathcal{N}(\mu, \sigma^2)$ and $U_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$ and the hyper-parameters in Eq. (3.17).

We now focus on the solution u at time $t = 1$. We thus define an abstract function f as

$$(3.18) \quad \begin{aligned} f: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (u_0, \lambda) &\mapsto F(u_0, \lambda, t = 1) = u_0 e^\lambda, \end{aligned}$$

and we define the output random variable of interest as $Y \equiv f(U_0, \Lambda)$. At $t = 1$, denoting $S_\star^1 \equiv S_\star(t = 1)$, the Sobol' indices defined above take the values

$$(3.19) \quad S_{U_0}^1 \approx 0.374, \quad S_\Lambda^1 \approx 0.602, \quad S_{U_0, \Lambda}^1 \approx 0.024,$$

and their numerators $V_\star^1 \equiv V_\star(t = 1) = S_\star^1 \times V(t = 1)$ are given by

$$(3.20) \quad V_{U_0}^1 \approx 0.577, \quad V_\Lambda^1 \approx 0.929, \quad V_{U_0, \Lambda}^1 \approx 0.037,$$

since $V(t = 1) \approx 1.543$. Furthermore, we define a sequence of abstract numerical simulators $(f_\ell)_{\ell \geq 0}$ as

$$(3.21) \quad \begin{aligned} f_\ell: \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (u_0, \lambda) &\mapsto u_\ell^{n_\ell}(u_0, \lambda), \end{aligned}$$

where

$$(3.22) \quad \begin{aligned} u_\ell^0: \mathbb{R}^2 &\rightarrow \mathbb{R}, \quad \text{and, } \forall k = 1, \dots, n_\ell, \quad u_\ell^k: \mathbb{R}^2 \rightarrow \mathbb{R}, \\ (u_0, \lambda) &\mapsto u_0 & (u_0, \lambda) &\mapsto \frac{u_\ell^{k-1}(u_0, \lambda)}{1 - \lambda/n_\ell}. \end{aligned}$$

In other words, f_ℓ maps a pair of values $(u_0, \lambda) \in \mathbb{R}^2$ to the corresponding numerical solution at time $t = 1$ using a backward Euler scheme with n_ℓ time steps. The discrete counterpart of Y on level ℓ is therefore $Y_\ell \equiv f_\ell(U_0, \Lambda)$.

We assume that evaluating f_ℓ for a given input pair (u_0, λ) requires $\mathcal{O}(n_\ell)$ operations. This corresponds to a realistic numerical time integration procedure where the solution at a given time step is computed from the solution at the previous time step, thus requiring to propagate the solution from the initial time $t = 0$ up to $t = 1$. In the following numerical experiments, we set $n_0 = 16$ equally spaced time steps on the coarsest level $\ell = 0$ to discretize the time interval $(0, 1]$, and we define $n_\ell = n_0 2^\ell$ as the number of time steps at level $\ell > 0$. Thus, n_ℓ clearly satisfies $n_\ell \approx s^\ell$ with $s = 2$ and the computational cost of f_ℓ is $\mathcal{O}(n_\ell^\gamma)$ with $\gamma = 1$.

The purpose of our numerical experiments is to study the advantage of MLMC sampling over standard MC sampling for the estimation of $V_{U_0}^1$ and V_Λ^1 , the numerators of the Sobol' indices $S_{U_0}^1$ and S_Λ^1 . The MLMC estimators $\hat{V}_{U_0}^1$ and \hat{V}_Λ^1 of $V_{U_0}^1$ and V_Λ^1 , respectively, are naturally defined by

$$(3.23) \quad \hat{V}_{U_0}^1 \equiv C_L^{\text{ML}}[f(U_0, \Lambda), f(U_0, \Lambda')] \quad \text{and} \quad \hat{V}_\Lambda^1 \equiv C_L^{\text{ML}}[f(U_0, \Lambda), f(U_0', \Lambda)],$$

for a given finest level L .

First, Fig. 2 shows the convergence properties of the numerical simulator when the number of time steps n_ℓ increases by a factor of $s = 2$ at each level ℓ . It allows to verify that the assumptions of Theorem 2.5 are satisfied before computing the MLMC estimators $\hat{V}_{U_0}^1$ and \hat{V}_Λ^1 . Based on the relaxed assumption for unbiased corrections, we see that the biases of the MLMC estimators $C_\ell^{\text{ML}}[f(U_0, \Lambda), f(U_0, \Lambda')]$ and $C_\ell^{\text{ML}}[f(U_0, \Lambda), f(U_0', \Lambda)]$ decrease in $\mathcal{O}(n_\ell^{-\alpha})$ with $\alpha = 1$. Furthermore, denoting $Y_\ell \equiv f_\ell(U_0, \Lambda)$, $Z_\ell \equiv f_\ell(U_0, \Lambda')$ and $\tilde{Z}_\ell \equiv f_\ell(U_0', \Lambda)$, we remark that $\mathbb{M}^4[\Delta_\ell^Y] = \mathbb{M}^4[\Delta_\ell^Z] = \mathbb{M}^4[\Delta_\ell^{\tilde{Z}}]$, so we conclude from Fig. 2 that

$$(3.24) \quad \mathbb{M}^4[\Delta_\ell^Y]^{1/2} + \mathbb{M}^4[\Delta_\ell^Z]^{1/2} = \mathbb{M}^4[\Delta_\ell^Y]^{1/2} + \mathbb{M}^4[\Delta_\ell^{\tilde{Z}}]^{1/2} = 2 \mathbb{M}^4[\Delta_\ell^Y]^{1/2}$$

decreases in $\mathcal{O}(n_\ell^{-\beta})$ with $\beta = 2$. Moreover, we note from the figure on the right-hand-side that $\{\mathbb{M}^4[Y_\ell]\}_{\ell \geq 0}$ is uniformly bounded, and thus so are $\{\mathbb{M}^4[Z_\ell]\}_{\ell \geq 0}$ and $\{\mathbb{M}^4[\tilde{Z}_\ell]\}_{\ell \geq 0}$. Consequently, all the assumptions of Theorem 2.5 hold, and thus the theorem applies for both $\hat{V}_{U_0}^1$ and \hat{V}_Λ^1 . In other words, given a target accuracy ε , there exists a discretization level L with $n_0 \times 2^L$ time steps and a sequence of sampling sizes $(M_\ell)_{\ell=0}^L$ such that the root mean square error of the MLMC estimator $\hat{V}_{U_0}^1$ (and, likewise, \hat{V}_Λ^1) is lower than this accuracy and its computational cost is $\mathcal{O}(\varepsilon^{-2})$.

On the other hand, the computational cost of standard Monte Carlo is $\mathcal{O}(\varepsilon^{-3})$, consisting in $\mathcal{O}(\varepsilon^{-2})$ simulations of the finest level f_L described by $\mathcal{O}(\varepsilon^{-1})$ time steps. From the reverse point of view, the accuracy measured in terms of root mean square error is $\mathcal{O}(C^{-1/2})$ for MLMC

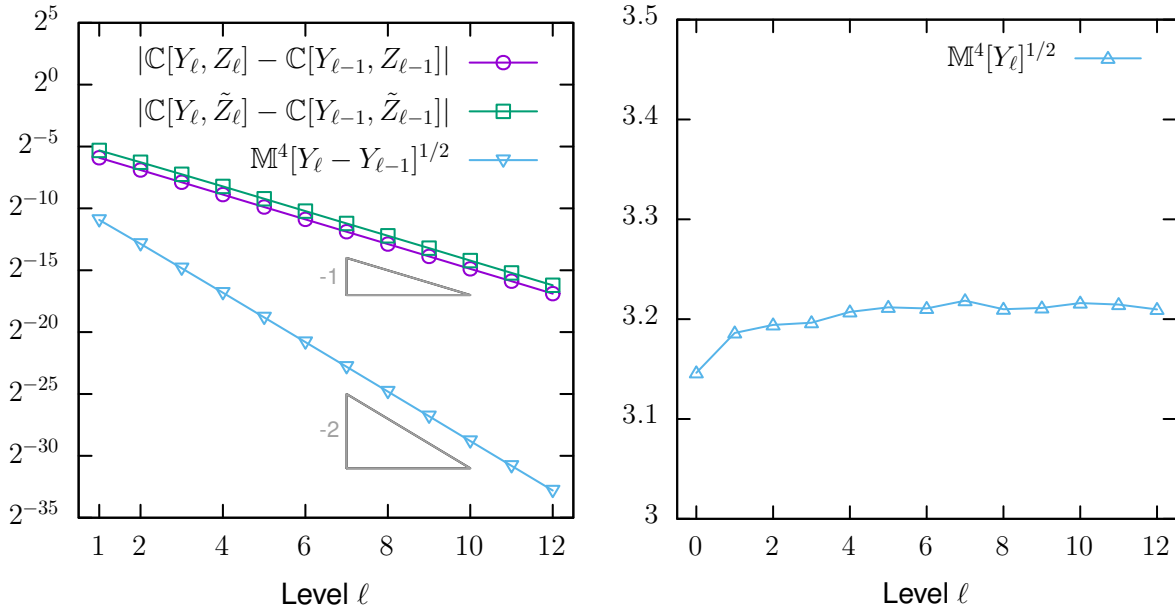


Figure 2: Convergence properties of the numerical simulator, where $Y_\ell \equiv f_\ell(U_0, \Lambda)$, $Z_\ell \equiv f_\ell(U_0, \Lambda')$ and $\tilde{Z}_\ell \equiv f_\ell(U'_0, \Lambda)$. We recall that \mathbb{M}^4 denotes the fourth-order central moment defined for any random variable $Y \in L_4(\Omega, \mathbb{P})$ as $\mathbb{M}^4[Y] \equiv \mathbb{E}[(Y - \mathbb{E}[Y])^4]$ (see section 2.4). The statistics are estimated using $M = 10^7$ samples on each level.

sampling and $\mathcal{O}(C^{-1/3})$ for standard MC sampling, where C denotes the computational cost of computing the estimator. These theoretical convergence rates of the RMSE as functions of the normalized cost C (see Remark 3.3 below) are numerically reflected in Fig. 3, based on 100 replications of the experiments. In particular, for the estimation of $V_{U_0}^1$ (Fig. 3a), we observe that with a normalized computational cost of $C \approx 3 \times 10^3$, the MLMC estimator achieves an accuracy of about 2×10^{-3} , while the MC estimator only has an accuracy of about 10^{-2} . Again, from the reverse point of view, using standard MC sampling to obtain an accuracy of 2×10^{-3} requires about 300 times more computational effort, *i.e.* a normalized cost $C \approx 10^6$, than when using MLMC. Similar observations can be made for the estimation of V_Λ^1 (Fig. 3b). To summarize, we see that multilevel Monte Carlo covariance estimators converge faster than the standard Monte Carlo ones, with a theoretical accuracy gain $\varepsilon_{\text{MC}}/\varepsilon_{\text{ML}}$ in $\mathcal{O}(C^{1/6})$, here numerically $\mathcal{O}(C^{0.21})$ and $\mathcal{O}(C^{0.29})$ for U_0 and Λ , respectively, and a theoretical speedup $C_{\text{MC}}/C_{\text{ML}}$ in $\mathcal{O}(\varepsilon^{-1})$.

Finally, Fig. 4 shows the evolution of the average sample size per level as a function of the total cost. For any given level ℓ , the sample size M_ℓ increases with the total cost. For the lower costs, only the first (coarsest) four levels out of a potential total of eleven are used. Then, as the total budget increases and coarsest corrections \hat{T}_ℓ have a small enough variance as compared to the discretization bias, finer levels are created and sampled in order to reduce the bias.

Remark 3.3. The cost is here arbitrarily normalized by the correction cost \mathcal{C}_8 on level

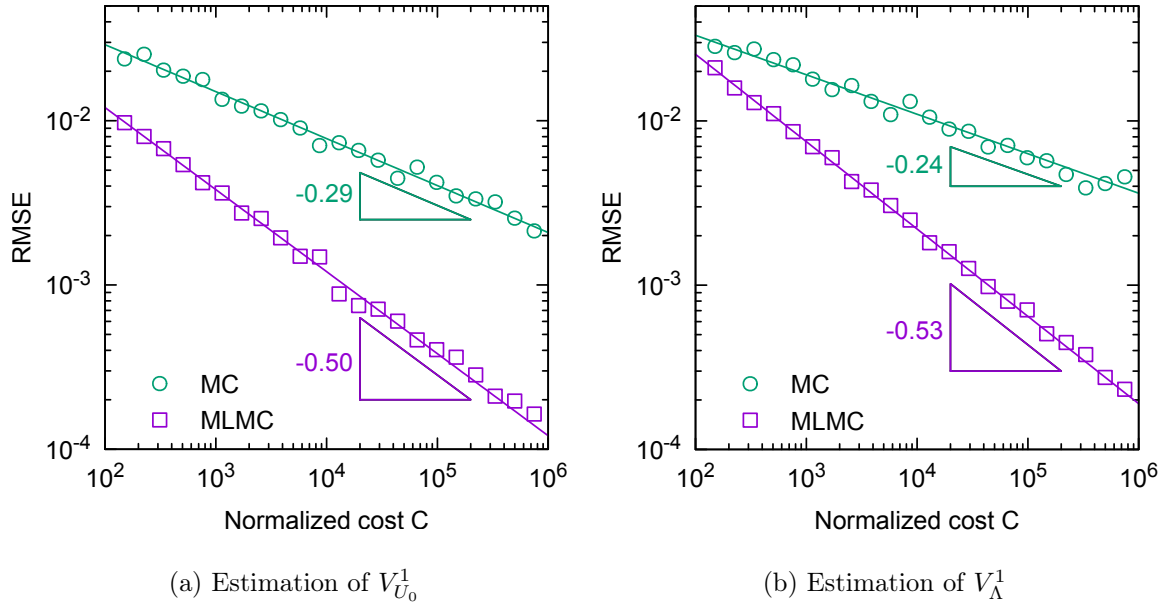


Figure 3: Evolution of the RMSE as a function of the normalized cost for the MC and MLMC estimation of $V_{U_0}^1$ (Fig. 3a) and V_{Λ}^1 (Fig. 3b), obtained by applying Algorithm 2 with $\tau = 1.5$ and $\rho = 0.25$. These statistics are estimated from 100 replications of the experiment.

$\ell = 8$. Specifically, because $C_\ell = 2^\ell C_0 = 2^{\ell-8} C_8$, we have

$$(3.25) \quad C \equiv C_8^{-1} \sum_{\ell=0}^L M_\ell C_\ell = \sum_{\ell=0}^L 2^{\ell-8} M_\ell.$$

Note that the normalized cost C only differs from the absolute cost $\sum_{\ell \leq L} M_\ell C_\ell$ by a fixed multiplicative factor, so that the convergence rates are unaffected by the normalization.

4. Conclusion. In computer experimentation, uncertainty quantification studies deal with the estimation of statistics (expectation, variance, quantile, probability, ...) or sensitivity indices associated to the output of a numerical simulator. The accuracy of crude Monte Carlo estimators is severely limited when the computer code is time-expensive, allowing only a small number of executions and leading to an important variance which is inversely proportional to the number of simulation runs. An alternative would be to replace the numerical simulator by a surrogate model, which could be evaluated at low cost during the Monte Carlo sampling, significantly reducing the sampling error but adding a model one. Another one would be to replace the Monte Carlo sampling by quasi Monte Carlo techniques in order to slightly reduce the estimator's variance.

In this paper driven by the estimation of Sobol' indices, we considered another alternative based on the Monte Carlo sampling of different simulator versions ranked by their exponentially

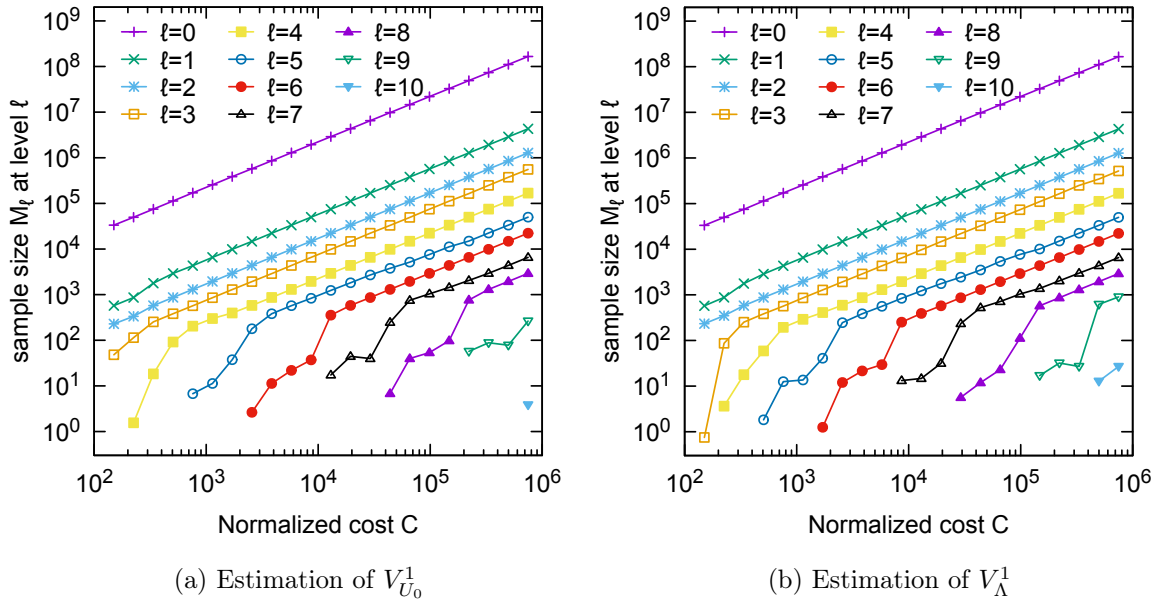


Figure 4: Evolution of the average sample size per level as a function of the normalized cost for the MC and MLMC estimation of $V_{U_0}^1$ (Fig. 4a) and V_{Λ}^1 (Fig. 4b), obtained by applying Algorithm 2 with $\tau = 1.5$ and $\rho = 0.25$. These statistics are estimated from 100 replications of the experiment.

increasing accuracy and computational cost. Precisely, we proposed a unified framework for the MLMC approach where the unbiased Monte Carlo estimator of the quantity of interest based on the finest level can be written as the telescoping sum of unbiased Monte Carlo estimators based on the lower levels. This framework can be used for different statistics with an unbiased estimator, such as the covariance present in the construction of a Sobol' index. We analyzed in particular the case of covariance estimation and we adapted the MLMC convergence theorem in terms of the corresponding covariances and fourth order moments. This theorem may be applied to the estimation of the covariance between two output instances, as well as between an output instance and the corresponding input. We also extended an existing iterative algorithm selecting the optimal number of runs at each level and the finest level by declining different optimal criteria. Whereas existing approaches typically consist in achieving a target accuracy for the estimator, our algorithm aims at complying with a given overall computational budget. This strategy may be more appropriate for engineering studies where one looks to reach the best accuracy under the constraint that the total simulation time is lower than a given requirement.

The proposed approach was efficiently studied and validated on an ordinary differential equation, for which a closed-form solution is available, with random parameters, namely a random initial condition and a random growth coefficient. In particular, we estimated covariance terms involved in the construction of the pick-and-freeze formulation of Sobol' indices, and

showed that MLMC had a better convergence rate than standard MC, which was confirmed numerically.

The results are promising for further MLMC developments dedicated to uncertainty quantification and management, such as sensitivity analysis with sensitivity indices based on output variance decomposition [49], output partial derivatives [10], output probability distribution [5] or other goal-oriented sensitivity indices [16]. It would also be appropriate to continue this work by proposing extensions of MLMC algorithms for the estimation of a group of sensitivity indices, along the same line as what was proposed in [20, Section 2.5] for multi-dimensional output functionals¹. Finally, further improvement may be obtained by combining the approach with QMC or randomized QMC sampling as in [24] or with the use of antithetic variates as in [22, 23].

Appendix A. Relaxed assumption for unbiased corrections. If $\hat{T}_{M_\ell}^{(\ell)}$ is an unbiased estimator of T_ℓ on each level $\ell \geq 0$, then

$$(A.1) \quad \text{Bias}(\hat{\theta}_\ell^{\text{ML}}, \theta) = \theta_\ell - \theta = - \sum_{k>\ell} \theta_k - \theta_{k-1}.$$

Assume $n_\ell \approx s^\ell$ for some fixed $s > 1$ and $|\theta_\ell - \theta_{\ell-1}| \lesssim n_\ell^{-\alpha}$, *i.e.* there exist positive constants a, b, c such that

$$(A.2) \quad as^\ell \leq n_\ell \leq bs^\ell \quad \text{and} \quad |\theta_\ell - \theta_{\ell-1}| \leq cn_\ell^{-\alpha}.$$

Then we have

$$(A.3) \quad |\text{Bias}(\hat{\theta}_\ell^{\text{ML}}, \theta)| \leq \sum_{k>\ell} |\theta_k - \theta_{k-1}| \leq a^{-\alpha} c \sum_{k>\ell} s^{-\alpha k} = a^{-\alpha} c \frac{s^{-\alpha \ell}}{s^\alpha - 1} \lesssim n_\ell^{-\alpha}.$$

Appendix B. Upper bound for the single-level MC estimators. Let $Y, Z \in L_4(\Omega, \mathbb{P})$. From Eq. (2.27) we notice that

$$(B.1) \quad \mathbb{V}[C_M[Y, Z]] = \frac{\mathbb{M}^4[Y, Z]}{M} + \frac{\mathbb{V}[Y] \mathbb{V}[Z]}{M(M-1)} - \frac{(M-2) \mathbb{C}[Y, Z]^2}{M(M-1)}$$

$$(B.2) \quad \leq \frac{\mathbb{M}^4[Y, Z]}{M} + \frac{\mathbb{V}[Y] \mathbb{V}[Z]}{M(M-1)}, \quad \forall M > 1.$$

Furthermore, from the Cauchy-Schwarz inequality, we have

$$(B.3) \quad \mathbb{M}^4[Y, Z] = \left| \langle \bar{Y}^2, \bar{Z}^2 \rangle_{L_2(\Omega, \mathbb{P})} \right| \leq \|\bar{Y}^2\|_{L_2(\Omega, \mathbb{P})} \|\bar{Z}^2\|_{L_2(\Omega, \mathbb{P})} = \sqrt{\mathbb{M}^4[Y] \mathbb{M}^4[Z]},$$

with the notation $\bar{U} \equiv U - \mathbb{E}[U]$ for any random variable U and using the fact that $U \in L_4(\Omega, \mathbb{P}) \implies \bar{U}^2 \in L_2(\Omega, \mathbb{P})$. Finally, using Jensen's inequality, we have

$$(B.4) \quad (\mathbb{V}[Y] \mathbb{V}[Z])^2 = \mathbb{E}[\bar{Y}^2]^2 \mathbb{E}[\bar{Z}^2]^2 \leq \mathbb{M}^4[Y] \mathbb{M}^4[Z],$$

¹attributed by the author to Tigran Nagapetyan.

leading to the bound in Eq. (2.29) for the covariance. The bound for the variance is obtained by taking $Y = Z$.

Appendix C. Upper bound for the multilevel covariance estimator.

Following the approach for the proof of [3, Theorem 5.1], we notice that

$$(C.1) \quad \mathbb{V}[C_L^{\text{ML}}[Y, Z]] = \sum_{\ell=0}^L \mathbb{V}[\hat{T}_{M_\ell}^{(\ell)}] = \sum_{\ell=0}^L \mathbb{E}\left[\left(\hat{T}_{M_\ell}^{(\ell)} - T_\ell\right)^2\right] = \sum_{\ell=0}^L \mathbb{E}[J_\ell^2],$$

where $J_\ell \equiv \hat{T}_{M_\ell}^{(\ell)} - T_\ell$ and

$$(C.2) \quad T_\ell \equiv \mathbb{C}[Y_\ell, Z_\ell] - \mathbb{C}[Y_{\ell-1}, Z_{\ell-1}], \quad \hat{T}_{M_\ell}^{(\ell)} \equiv C_{M_\ell}^{(\ell)}[Y_\ell, Z_\ell] - C_{M_\ell}^{(\ell)}[Y_{\ell-1}, Z_{\ell-1}].$$

Furthermore, we have the identities

$$(C.3) \quad \mathbb{C}[Y_\ell, Z_\ell] - \mathbb{C}[Y_{\ell-1}, Z_{\ell-1}] = \mathbb{C}[\Delta_\ell^Y, Z_\ell] + \mathbb{C}[Y_{\ell-1}, \Delta_\ell^Z]$$

$$(C.4) \quad = \mathbb{C}[\Delta_\ell^Y, Z_{\ell-1}] + \mathbb{C}[Y_\ell, \Delta_\ell^Z]$$

$$(C.5) \quad = \frac{1}{2} (\mathbb{C}[\Delta_\ell^Y, \Sigma_\ell^Z] + \mathbb{C}[\Sigma_\ell^Y, \Delta_\ell^Z]),$$

and, likewise,

$$(C.6) \quad C_{M_\ell}^{(\ell)}[Y_\ell, Z_\ell] - C_{M_\ell}^{(\ell)}[Y_{\ell-1}, Z_{\ell-1}] = \frac{1}{2} \left(C_{M_\ell}^{(\ell)}[\Delta_\ell^Y, \Sigma_\ell^Z] + C_{M_\ell}^{(\ell)}[\Sigma_\ell^Y, \Delta_\ell^Z] \right).$$

Thus, we have

$$(C.7) \quad J_\ell^2 = \frac{1}{4} (A_\ell + B_\ell)^2 \leq \frac{1}{2} (A_\ell^2 + B_\ell^2),$$

with

$$(C.8) \quad A_\ell \equiv C_{M_\ell}^{(\ell)}[\Delta_\ell^Y, \Sigma_\ell^Z] - \mathbb{C}[\Delta_\ell^Y, \Sigma_\ell^Z], \quad B_\ell \equiv C_{M_\ell}^{(\ell)}[\Sigma_\ell^Y, \Delta_\ell^Z] - \mathbb{C}[\Sigma_\ell^Y, \Delta_\ell^Z].$$

Using the bound in Eq. (2.29) for $\mathbb{E}[A_\ell^2] = \mathbb{V}[C_{M_\ell}^{(\ell)}[\Delta_\ell^Y, \Sigma_\ell^Z]]$ on the one hand, and for $\mathbb{E}[B_\ell^2] = \mathbb{V}[C_{M_\ell}^{(\ell)}[\Sigma_\ell^Y, \Delta_\ell^Z]]$ on the other hand, we obtain Eq. (2.31).

Furthermore, suppose that $\{\mathbb{M}^4[Y_\ell]\}_{\ell \geq 0}$ and $\{\mathbb{M}^4[Z_\ell]\}_{\ell \geq 0}$ are uniformly bounded, *i.e.* there exist constants $c_Y \geq 0$ and $c_Z \geq 0$, independent from any parameters, such that

$$(C.9) \quad \forall \ell \geq 0, \quad \mathbb{M}^4[Y_\ell] \leq c_Y \quad \text{and} \quad \mathbb{M}^4[Z_\ell] \leq c_Z.$$

We then have

$$(C.10) \quad \mathbb{M}^4[\Sigma_\ell^Y] = \|(\bar{Y}_\ell + \bar{Y}_{\ell-1})^2\|_{L_2(\Omega, \mathbb{P})}^2 \leq 4 \|\bar{Y}_\ell^2 + \bar{Y}_{\ell-1}^2\|_{L_2(\Omega, \mathbb{P})}^2$$

$$(C.11) \quad \leq 4 \left(\mathbb{M}^4[Y_\ell]^{1/2} + \mathbb{M}^4[Y_{\ell-1}]^{1/2} \right)^2 \leq 16c_Y.$$

The same naturally holds for Z , *i.e.* $\mathbb{M}^4[\Sigma_\ell^Z] \leq 16c_Z$. Denoting $c \equiv 2 \max(c_Y, c_Z)^{1/2}$, it follows that

$$(C.12) \quad \mathbb{V}[C_L^{\text{ML}}[Y, Z]] \leq c \sum_{\ell \leq L} \frac{1}{M_\ell - 1} \left[\mathbb{M}^4[\Delta_\ell^Y]^{1/2} + \mathbb{M}^4[\Delta_\ell^Z]^{1/2} \right].$$

Appendix D. Statistics of the solution of the initial value problem.

We provide here the detailed derivation of Eqs. (3.11) to (3.16). By definition, and because U_0 , U'_0 , Λ and Λ' are independent, we have

$$(D.1) \quad V_{U_0}(t) = \mathbb{C}[F(U_0, \Lambda, t), F(U_0, \Lambda', t)] = \mathbb{C}[U_0 e^{\Lambda t}, U_0 e^{\Lambda' t}]$$

$$(D.2) \quad = \mathbb{E}[U_0^2 e^{\Lambda t} e^{\Lambda' t}] - \mathbb{E}[U_0 e^{\Lambda t}] \mathbb{E}[U_0 e^{\Lambda' t}] = \mathbb{E}[U_0^2] \mathbb{E}[e^{\Lambda t}]^2 - \mathbb{E}[U_0]^2 \mathbb{E}[e^{\Lambda t}]^2$$

$$(D.3) \quad = \mathbb{E}[e^{\Lambda t}]^2 (\mathbb{E}[U_0^2] - \mathbb{E}[U_0]^2) = \mathbb{E}[e^{\Lambda t}]^2 \mathbb{V}[U_0].$$

Likewise,

$$(D.4) \quad V_\Lambda(t) = \mathbb{C}[U_0 e^{\Lambda t}, U'_0 e^{\Lambda t}] = \mathbb{E}[U_0^2] \mathbb{E}[e^{2\Lambda t}] - \mathbb{E}[U_0]^2 \mathbb{E}[e^{\Lambda t}]^2 = \mathbb{E}[U_0^2] \mathbb{V}[e^{\Lambda t}],$$

$$(D.5) \quad V(t) = \mathbb{E}[U_0^2 e^{2\Lambda t}] - \mathbb{E}[U_0 e^{\Lambda t}]^2 = \mathbb{E}[U_0^2] \mathbb{E}[e^{2\Lambda t}] - \mathbb{E}[U_0]^2 \mathbb{E}[e^{\Lambda t}]^2.$$

Then, for the particular case where $\Lambda \sim \mathcal{N}(\mu, \sigma^2)$ and $U_0 \sim \mathcal{N}(\mu_0, \sigma_0^2)$, it is easy to show that

$$(D.6) \quad \mathbb{E}[U_0^2] = \sigma_0^2 + \mu_0^2, \quad \mathbb{E}[e^{\Lambda t}] = e^{\mu t + \frac{\sigma^2 t^2}{2}},$$

$$(D.7) \quad \mathbb{E}[e^{2\Lambda t}] = e^{2(\mu t + \sigma^2 t^2)}, \quad \mathbb{V}[e^{\Lambda t}] = (e^{\sigma^2 t^2} - 1) e^{2\mu t + \sigma^2 t^2},$$

leading to Eqs. (3.14), (3.15) and (3.16).

Acknowledgments. The authors are grateful to Gustavo Vila for his careful proofreading of the manuscript resulting in the correction of several typographical errors.

REFERENCES

- [1] L. J. ASLETT, T. NAGAPETYAN, AND S. J. VOLLMER, *Multilevel Monte Carlo for reliability theory*, Reliability Engineering & System Safety, 165 (2017), pp. 188–196, <https://doi.org/10.1016/j.res.2017.03.003>, <http://www.sciencedirect.com/science/article/pii/S0951832016302551>.
- [2] S. BARTHÉLÉMY, S. RICCI, E. LE PAPE, M. ROCHOUX, O. THUAL, N. GOUTAL, J. HABERT, A. PIA-CENTINI, G. JONVILLE, F. ZAOUI, AND P. GOUIN, *Ensemble-based algorithm for error reduction in hydraulics in the context of flood forecasting*, FLOODrisk 2016 Conference, 2016.
- [3] C. BIERIG AND A. CHERNOV, *Convergence analysis of multilevel Monte Carlo variance estimators and application for random obstacle problems*, Numerische Mathematik, 130 (2015), pp. 579–613, <https://doi.org/10.1007/s00211-014-0676-3>, <https://doi.org/10.1007/s00211-014-0676-3>.
- [4] C. BIERIG AND A. CHERNOV, *Estimation of arbitrary order central statistical moments by the multilevel Monte Carlo method*, Stochastics and Partial Differential Equations Analysis and Computations, 4 (2016), pp. 3–40, <https://doi.org/10.1007/s40072-015-0063-9>, <https://doi.org/10.1007/s40072-015-0063-9>.
- [5] E. BORGONOVO, *A new uncertainty importance measure*, Reliability Engineering & System Safety, 92 (2007), pp. 771–784, <https://doi.org/https://doi.org/10.1016/j.res.2006.04.015>, <http://www.sciencedirect.com/science/article/pii/S0951832006000883>.

- [6] R. E. CAFLISCH, *Monte Carlo and quasi-Monte Carlo methods*, Acta Numerica, 7 (1998), pp. 1–49, <https://doi.org/10.1017/S0962492900002804>, <https://doi.org/10.1017/S0962492900002804>.
- [7] K. A. CLIFFE, M. B. GILES, R. SCHEICHL, AND A. L. TECKENTRUP, *Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients*, Computing and Visualization in Science, 14 (2011), pp. 3–15, <https://doi.org/10.1007/s00791-011-0160-x>, <http://dx.doi.org/10.1007/s00791-011-0160-x>.
- [8] N. COLLIER, A.-L. HAJI-ALI, F. NOBILE, E. VON SCHWERIN, AND R. TEMPONE, *A continuation multilevel Monte Carlo algorithm*, BIT Numerical Mathematics, 55 (2015), pp. 399–432, <https://doi.org/10.1007/s10543-014-0511-3>, <https://doi.org/10.1007/s10543-014-0511-3>.
- [9] T. CRESTAUX, O. LE MAÎTRE, AND J.-M. MARTINEZ, *Polynomial chaos expansion for sensitivity analysis*, Reliability Engineering & System Safety, 94 (2009), pp. 1161–1172, <https://doi.org/https://doi.org/10.1016/j.res.2008.10.008>, <http://www.sciencedirect.com/science/article/pii/S0951832008002561>. Special Issue on Sensitivity Analysis.
- [10] M. DE LOZZO AND A. MARREL, *Estimation of the derivative-based global sensitivity measures using a Gaussian process metamodel*, SIAM/ASA Journal on Uncertainty Quantification, 4 (2016), pp. 708–738, <https://doi.org/10.1137/15M1013377>, <https://doi.org/10.1137/15M1013377>, <https://arxiv.org/abs/https://doi.org/10.1137/15M1013377>.
- [11] E. DE ROCQUIGNY, N. DEVICTOR, AND S. TARANTOLA, *Uncertainty in Industrial Practice*, John Wiley & Sons, Ltd, 1 ed., 2008, <https://doi.org/10.1002/9780470770733>, <https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470770733>, <https://arxiv.org/abs/https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470770733>.
- [12] S. DOLGOV, B. N. KHOROMSKIJ, A. LITVINENKO, AND H. G. MATTHIES, *Computation of the response surface in the tensor train data format*, arXiv preprint arXiv:1406.2816, (2014), <https://arxiv.org/abs/1406.2816>.
- [13] B. EFRON AND C. STEIN, *The jackknife estimate of variance*, The Annals of Statistics, 9 (1981), pp. 586–596, <https://doi.org/10.1214/aos/1176345462>, <https://doi.org/10.1214/aos/1176345462>.
- [14] M. ESPIG, W. HACKBUSCH, A. LITVINENKO, H. G. MATTHIES, AND P. WÄHNERT, *Efficient low-rank approximation of the stochastic Galerkin matrix in tensor formats*, Computers & Mathematics with Applications, 67 (2014), pp. 818–829, <https://doi.org/https://doi.org/10.1016/j.camwa.2012.10.008>, <http://www.sciencedirect.com/science/article/pii/S0898122112006244>. High-order Finite Element Approximation for Partial Differential Equations.
- [15] A. FORRESTER, A. SOBESTER, AND A. KEANE, *Engineering Design via Surrogate Modelling: A Practical Guide*, Wiley, 2008, <https://doi.org/10.1002/9780470770801>, <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470770801>.
- [16] J.-C. FORT, T. KLEIN, AND N. RACHDI, *New sensitivity analysis subordinated to a contrast*, Communications in Statistics - Theory and Methods, 45 (2016), pp. 4349–4364, <https://doi.org/10.1080/03610926.2014.901369>, <https://doi.org/10.1080/03610926.2014.901369>, <https://arxiv.org/abs/https://doi.org/10.1080/03610926.2014.901369>.
- [17] R. GHANEM, R. HIGDON, AND H. OWHADI, eds., *Handbook of Uncertainty Quantification*, Springer International Publishing, 2017, <https://doi.org/10.1007/978-3-319-12385-1>, <https://doi.org/10.1007/978-3-319-12385-1>.
- [18] R. GHANEM AND P. SPANOS, *Stochastic Finite Elements: A Spectral Approach*, Civil, Mechanical and Other Engineering Series, Dover Publications, 2003, <https://doi.org/10.1007/978-1-4612-3094-6>, <https://www.springer.com/us/book/9781461277958>.
- [19] M. GILES, *Multilevel Monte Carlo path simulation*, Operations Research, 3 (2008), pp. 607–617.
- [20] M. GILES, *Multilevel Monte Carlo methods*, Acta Numerica, 24 (2015), pp. 259–328, <https://doi.org/10.1017/S096249291500001X>, <https://doi.org/10.1017/S096249291500001X>.
- [21] M. GILES AND L. SZPRUCH, *Multilevel Monte Carlo methods for applications in finance*, World Scientific, 2013, pp. 3–47, https://doi.org/10.1142/9789814436434_0001, https://www.worldscientific.com/doi/abs/10.1142/9789814436434_0001, https://arxiv.org/abs/https://www.worldscientific.com/doi/pdf/10.1142/9789814436434_0001.
- [22] M. B. GILES AND L. SZPRUCH, *Antithetic multilevel Monte Carlo estimation for multidimensional SDEs*, in Monte Carlo and Quasi-Monte Carlo Methods 2012, J. Dick, F. Y. Kuo, G. W. Peters, and I. H. Sloan, eds., Berlin, Heidelberg, 2013, Springer Berlin Heidelberg, pp. 367–384, https://doi.org/10.1007/978-3-642-30944-6_17.

- 1007/978-3-642-41095-6_16, https://doi.org/10.1007/978-3-642-41095-6_16.
- [23] M. B. GILES AND L. SZPRUCH, *Antithetic multilevel Monte Carlo estimation for multi-dimensional SDEs without Lévy area simulation*, The Annals of Applied Probability, 24 (2014), pp. 1585–1620, <https://doi.org/10.1214/13-AAP957>, <https://doi.org/10.1214/13-AAP957>.
- [24] M. B. GILES AND B. J. WATERHOUSE, *Multilevel quasi-Monte Carlo path simulation*, Advanced Financial Modelling, Radon Series on Computational and Applied Mathematics, (2009), pp. 165–181.
- [25] J. HAJEK, *Asymptotic normality of simple linear rank statistics under alternatives*, Ann. Math. Statist., 39 (1968), pp. 325–346, <https://doi.org/10.1214/aoms/1177698394>, <https://doi.org/10.1214/aoms/1177698394>.
- [26] A.-L. HAJI-ALI, F. NOBILE, AND R. TEMPONE, *Multi-index Monte Carlo: when sparsity meets sampling*, Numerische Mathematik, 132 (2016), pp. 767–806, <https://doi.org/10.1007/s00211-015-0734-5>, <https://doi.org/10.1007/s00211-015-0734-5>.
- [27] A.-L. HAJI-ALI, F. NOBILE, E. VON SCHWERIN, AND R. TEMPONE, *Optimization of mesh hierarchies in multilevel Monte Carlo samplers*, Stochastics and Partial Differential Equations Analysis and Computations, 4 (2016), pp. 76–112, <https://doi.org/10.1007/s40072-015-0049-7>, <https://doi.org/10.1007/s40072-015-0049-7>.
- [28] S. HEINRICH, *Monte Carlo complexity of global solution of integral equations*, Journal of Complexity, 14 (1998), pp. 151–175, <https://doi.org/https://doi.org/10.1006/jcom.1998.0471>, <http://www.sciencedirect.com/science/article/pii/S0885064X9890471X>.
- [29] S. HEINRICH, *Multilevel Monte Carlo methods*, in Large-Scale Scientific Computing, S. Margenov, J. Waśniewski, and P. Yalamov, eds., Berlin, Heidelberg, 2001, Springer Berlin Heidelberg, pp. 58–67, https://doi.org/10.1007/3-540-45346-6_5, https://doi.org/10.1007/3-540-45346-6_5.
- [30] S. HEINRICH AND E. SINDAMBIWE, *Monte Carlo complexity of parametric integration*, Journal of Complexity, 15 (1999), pp. 317–341, <https://doi.org/https://doi.org/10.1006/jcom.1999.0508>, <http://www.sciencedirect.com/science/article/pii/S0885064X99905083>.
- [31] W. HOEFFDING, *A class of statistics with asymptotically normal distribution*, The Annals of Mathematical Statistics, 19 (1948), pp. 293–325, <https://doi.org/10.1214/aoms/1177730196>, <https://doi.org/10.1214/aoms/1177730196>.
- [32] H. HOEL, K. LAW, AND R. TEMPONE, *Multilevel ensemble Kalman filtering*, SIAM Journal on Numerical Analysis, 54 (2016), pp. 1813–1839, <https://doi.org/10.1137/15M100955X>, <https://doi.org/10.1137/15M100955X>, <https://arxiv.org/abs/https://doi.org/10.1137/15M100955X>.
- [33] A. JANON, *Sensitivity analysis and model reduction : application to oceanography*, PhD thesis, Université de Grenoble, Nov. 2012, <https://tel.archives-ouvertes.fr/tel-00757101>.
- [34] A. JANON, T. KLEIN, A. LAGNOUX, M. NODET, AND C. PRIEUR, *Asymptotic normality and efficiency of two Sobol index estimators*, ESAIM: PS, 18 (2014), pp. 342–364, <https://doi.org/10.1051/ps/2013040>, <https://doi.org/10.1051/ps/2013040>.
- [35] G. KATSIOLIDES, E. H. MÜLLER, R. SCHEICHL, T. SHARDLOW, M. B. GILES, AND D. J. THOMSON, *Multilevel Monte Carlo and improved timestepping methods in atmospheric dispersion modelling*, Journal of Computational Physics, 354 (2018), pp. 320–343, <https://doi.org/https://doi.org/10.1016/j.jcp.2017.10.035>, <http://www.sciencedirect.com/science/article/pii/S0021999117307945>.
- [36] J. P. KLEIJNEN, *Sensitivity analysis and related analyses: A review of some statistical techniques*, Journal of Statistical Computation and Simulation, 57 (1997), pp. 111–142, <https://doi.org/10.1080/00949659708811805>, <https://doi.org/10.1080/00949659708811805>, <https://arxiv.org/abs/https://doi.org/10.1080/00949659708811805>.
- [37] F. KUO, R. SCHEICHL, C. SCHWAB, I. SLOAN, AND E. ULLMANN, *Multilevel quasi-Monte Carlo methods for lognormal diffusion problems*, Mathematics of Computation, 86 (2017), pp. 2827–2860, <https://doi.org/10.1090/mcom/3207>, <https://doi.org/10.1090/mcom/3207>.
- [38] J. LATZ, I. PAPAIOANNOU, AND E. ULLMANN, *Multilevel sequential Monte Carlo for Bayesian inverse problems*, Journal of Computational Physics, 368 (2018), pp. 154–178, <https://doi.org/https://doi.org/10.1016/j.jcp.2018.04.014>, <http://www.sciencedirect.com/science/article/pii/S0021999118302286>.
- [39] O. LE MAÎTRE AND O. M. KNIO, *Spectral methods for uncertainty quantification: with applications to computational fluid dynamics*, Springer Netherlands, 2010, <https://doi.org/10.1007/978-90-481-3520-2>, <https://doi.org/10.1007/978-90-481-3520-2>.

- [40] C. LEMIEUX, *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer, 2010, <https://doi.org/10.1007/978-0-387-78165-5>.
- [41] C. LESTER, R. E. BAKER, M. B. GILES, AND C. A. YATES, *Extending the multi-level method for the simulation of stochastic biological systems*, *Bulletin of Mathematical Biology*, 78 (2016), pp. 1640–1677, <https://doi.org/10.1007/s11538-016-0178-9>, <https://doi.org/10.1007/s11538-016-0178-9>.
- [42] A. NOUY, *Proper generalized decompositions and separated representations for the numerical solution of high dimensional stochastic problems*, *Archives of Computational Methods in Engineering*, 17 (2010), pp. 403–434, <https://doi.org/10.1007/s11831-010-9054-1>, <https://doi.org/10.1007/s11831-010-9054-1>.
- [43] C. E. RASMUSSEN AND C. K. I. WILLIAMS, *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*, The MIT Press, 2005.
- [44] C. ROBERT AND G. CASELLA, *Monte Carlo Statistical Methods*, Springer-Verlag, 2010, <https://doi.org/10.1007/978-1-4757-4145-2>, <https://doi.org/10.1007/978-1-4757-4145-2>.
- [45] I. ROSENBAUM AND J. STAUM, *Multilevel Monte Carlo metamodeling*, *Operations Research*, 65 (2017), pp. 1062–1077, <https://doi.org/10.1287/opre.2017.1607>, <https://doi.org/10.1287/opre.2017.1607>, <https://arxiv.org/abs/https://doi.org/10.1287/opre.2017.1607>.
- [46] A. SALTELLI, K. CHAN, AND E. SCOTT, *Sensitivity analysis*, *Wiley Series in Probability and Statistics*, John Wiley & Sons, Ltd., Chichester, 2008.
- [47] T. J. SANTNER, B. J. WILLIAMS, AND W. NOTZ, *The design and analysis of computer experiments*, Springer-Verlag, 2003, <https://doi.org/10.1007/978-1-4757-3799-8>, <https://doi.org/10.1007/978-1-4757-3799-8>.
- [48] R. SMITH, *Uncertainty Quantification: Theory, Implementation, and Applications*, *Computational Science and Engineering*, SIAM, 2014.
- [49] I. SOBOL, *Sensitivity estimates for nonlinear mathematical models*, *Math. Modeling Comput. Experiment*, 4 (1993), pp. 407–414.
- [50] B. SUDRET, *Global sensitivity analysis using polynomial chaos expansions*, *Reliability Engineering & System Safety*, 93 (2008), pp. 964–979, <https://doi.org/https://doi.org/10.1016/j.res.2007.04.002>, <http://www.sciencedirect.com/science/article/pii/S0951832007001329>. Bayesian Networks in Dependability.
- [51] A. TECKENTRUP, *Multilevel Monte Carlo Methods and Uncertainty Quantification*, PhD thesis, University of Bath, 6 2013.
- [52] A. W. VAN DER VAART, *Asymptotic Statistics*, *Cambridge Series in Statistical and Probabilistic Mathematics*, Cambridge University Press, 1998, <https://doi.org/10.1017/CBO9780511802256>.
- [53] D. XIU, *Numerical methods for stochastic computations: A spectral method approach*, Princeton University Press, United States, 7 2010.